

Using Massive Online Choice Experiments to Measure Changes in Well-being

Erik Brynjolfsson
Massachusetts Institute of Technology and NBER

Felix Eggers
University of Groningen

Avinash Gannamaneni
Massachusetts Institute of Technology

October 2017

Abstract

GDP and metrics derived from it, like productivity, have been central to our understanding of economic progress and well-being. In principle, changes in consumer surplus (compensating expenditure) provide a superior, and more direct, measure of changes in consumer well-being, especially for digital goods. In practice, consumer surplus has been difficult to measure. We explore the potential of online Single Binary Discrete Choice (SBDC) experiments that seek to measure consumers' willingness to accept compensation for losing access to various digital goods and thereby estimate the changes in consumer surplus from these goods. We test the robustness of the approach and benchmark it against established methods, including a comparison of hypothetical choice and incentive compatible choice that require participants to give up Facebook for a period of time in exchange for compensation. The proposed choice experiments show convergent validity and are massively scalable. Our results indicate that digital goods have created enormous gains in well-being which are largely missed by conventional measures of GDP and productivity. By periodically querying a large, representative sample of goods and services, including those which are not priced in existing markets, changes in consumer surplus and other new measures of well-being derived from these online choice experiments have the potential for providing cost-effective supplements to existing national income and product accounts.

Keywords: Consumer Surplus, Digital Goods, Free Goods, GDP, Choice Experiments, Well-being

JEL Classification: C82, I30, O40

Acknowledgements: We thank Susanto Basu, Erwin Diewert, Robert Hall, John Hauser, Leonard Nakamura, Jeff Prince, Adam Saunders, Christopher Stanton and participants of the NBER Summer workshop for the Conference on Research on Income and Wealth (CRIW 2016), NBER Winter Digitization meeting (2017) and NBER Productivity lunch (2017) for helpful comments and the MIT Initiative on the Digital Economy, through a grant from the Markle Foundation, for generous funding.

“If you don't know where you're going, you might not get there.” -- Yogi Berra

1. Introduction

Digital technologies have transformed the nature of production and the types of goods and services consumed in modern economies. Yet our measurement framework for economic growth and well-being has not fundamentally changed since the 1930s. In principle, a more comprehensive approach is now feasible. By using massive online choice experiments to estimate changes in consumer surplus (compensating variation) we can supplement the traditional metrics based on Gross Domestic Product (GDP).

GDP measures the monetary value of the purchases of all final goods by households, businesses and government. It is the most widely used measure of economic activity and heavily influences policymakers in setting economic objectives. GDP has been heralded as one of the greatest inventions of the 20th century by Paul Samuelson and William Nordhaus (Landefeld 2000). Both economists and journalists routinely use GDP as if it were a welfare measure. Media articles regularly mention that the “economy grew by x%”¹ by measuring the growth in GDP, and use this figure as a casual metric for the improvement in economic well-being. Similarly, economists widely use GDP per hour worked as a measure of productivity and infer links between productivity and improvement in living standards (OECD 2008).

However many economists consider GDP to be a significantly flawed measure of well-being and several attempts have been made to design alternative measures (Stiglitz et al. 2009). In fact, Simon Kuznets, the founding father of the system of national accounts that include GDP,

¹ E.g. U.S. Economy Grew 1.4% in Fourth Quarter (<http://www.bloomberg.com/news/articles/2016-03-25/u-s-economy-grew-1-4-in-fourth-quarter-supported-by-consumers>), China's Economy Grew by 6.7% in First Quarter of 2016 (<http://blogs.wsj.com/chinarealtime/2016/04/15/chinas-economy-grew-by-6-7-in-first-quarter-of-2016/>)

explicitly warned against using it this way, writing “The welfare of a nation can scarcely be inferred from a measurement of national income as defined [by the GDP.]” (Kuznets 1934).² Despite Kuznets’ warning, growth in GDP is still the most widely used indicator of progress in our economic well-being.

For goods with a non-zero price, in theory it is often possible to infer welfare from national accounts including GDP measures (Hulten 1978, Diewert 1983, Jorgenson and Slesnick 2014), although in practice official estimates of welfare are not published. Research has looked at factors such as introduction of new goods, intangibles, quality adjustments and household production when GDP is biased away from welfare and ways to correct these biases have been proposed.

GDP as a welfare measure is especially problematic in the emerging digital economy since most of the digital goods have nearly zero marginal cost and often a zero equilibrium price. This makes it difficult to discern their contributions by looking at GDP calculations (Brynjolfsson and Saunders 2009; Brynjolfsson and McAfee 2014). For instance, although information goods have become increasingly ubiquitous and important in our daily lives, the share of the information sector as a fraction of the total GDP (~ 4-5%) has not changed in the last 35 years (Figure 1). Moreover, in many sectors (e.g. music, media, encyclopedias) people substitute paid goods with zero-price online services (e.g., Spotify, YouTube, Wikipedia) so that the total revenue that shows up in GDP figures could fall even while consumers get access to better quality and more variety of digital goods (Brynjolfsson and Saunders 2009). In other words, not only the magnitude, but even the sign of the change in well-being may be incorrectly

² He underscored his views when accepting his Nobel Prize in 1971, saying that the conventional measures of national product (including GDP) omit various costs (e.g. pollution) and benefits (e.g. more leisure time) associated with technological innovations and predicted major changes in the way we measure the economy (Kuznets 1973).

inferred if decision makers rely solely on existing measures of GDP and productivity as a proxy for well-being.

[Insert Figure 1 here]

The benefits of technological advance are distinct from the expenditures on goods and services. Nordhaus (2005) estimated that between 1948 and 2001 corporations were able to retain only 3.7% of the social returns from their technological advances while the remaining 96.3% of social returns went to consumers. Consumer surplus thus reflects most of the returns to improvements in technology.

So far, the change in consumer surplus hasn't been widely used as a measure of change in well-being not because it is a poor measure of well-being, but because it is difficult to measure at scale. Estimating demand curves using market data requires exogenous variations that shift the supply curve but not the demand curve and it has not been practical to identify these variations for large bundles of goods.

However, with advances in digital technologies it is now feasible to collect data about thousands of goods easily, potentially improving our metrics. These big data techniques have the potential to improve measurement of economic indicators. For example, Cavallo and Rigobon (2016) scrape the web to collect billions of prices for millions of products to construct prices indexes and inflation measures for various countries as part of MIT's Billion Prices Project. Private companies like Microsoft, Amazon, Google and Facebook routinely conduct millions of online experiments to help understand consumer preferences and behavior. This scale of experimentation and inference would have been infeasible 20 years ago, but is now routine at many organizations.

In this research, we propose a way of measuring the changes in consumer surplus, not only for goods and services in the digital economy but also more broadly. Specifically, we implement a series of discrete choice experiments that measure consumers' willingness to accept payment in exchange for losing access to various goods. These experiments allow us to estimate the demand curves for these goods using data from thousands of consumers that are representative of the US population. We conclude that our approach is easily scalable and can be used to develop a system that tracks changes in consumer surplus of numerous goods and services in (near) real time via massive online choice experiments.

The paper proceeds as follows. In section 2, we illustrate the ways the GDP and consumer surplus change when prices change or new products are introduced, and the implications for welfare estimates. Section 3 describes the key methodologies we use to empirically assess consumer surplus. Section 4 provides results and sensitivity analyses of the proposed method. Section 5 applies the method to a broader set of goods. Section 6 concludes with a summary and discussion.

2. Background

2.1 GDP, consumer surplus and well-being

Perhaps no one has described the shortcomings of GDP³ as a welfare measure as eloquently as Robert F. Kennedy:

Gross National Product counts air pollution and cigarette advertising, and ambulances to clear our highways of carnage. It counts special locks for our doors and the jails for

³ Kennedy was technically discussing GNP, but his comments are equally applicable to GDP.

the people who break them. It counts napalm and counts nuclear warheads and armored cars for the police to fight the riots in our cities...

Yet the gross national product does not allow for the health of our children, the quality of their education or the joy of their play. It does not include the beauty of our poetry or the strength of our marriages, the intelligence of our public debate or the integrity of our public officials.

It measures neither our wit nor our courage, neither our wisdom nor our learning, neither our compassion nor our devotion to our country, it measures everything in short, except that which makes life worthwhile.⁴

Kennedy's poetic words contribute much to our understanding (if not to our GDP!) and there have been a number of efforts to create a more comprehensive estimate of well-being. Since 2012, the United Nations Sustainable Development Solutions Network published an annual World Happiness Report ranking countries based on measures of happiness (Helliwell et. al. 2017). Jones and Klenow (2016) propose a measure that incorporates consumption, leisure, mortality and inequality to measure the economic well-being of a country. There is a growing stream of literature focusing on measuring subjective well-being and life satisfaction. However, a survey of leading macroeconomists indicates that we are a long way off from reaching consensus on how to measure well-being so that they are reliable for policymaking (den Haan et. al. 2017).

In this paper, we are less ambitious and seek to stick more closely to a traditional microeconomic framework. In particular, we focus on the changes in consumer surplus generated by digital goods and discuss ways in which our approach can be expanded to more goods and services. Brynjolfsson and Saunders (2009) paraphrase Robert Solow in noting that the influence

⁴ Robert Kennedy speaking at University of Kansas in 1968 (Ref: <http://www.jfklibrary.org/Research/Research-Aids/Ready-Reference/RFK-Speeches/Remarks-of-Robert-F-Kennedy-at-the-University-of-Kansas-March-18-1968.aspx>).

of the information age is seen everywhere except in the GDP statistics. Almost all of us use more and more digital goods such as search engines, smartphones, social networking sites, e-commerce platforms but their revenues don't always reflect this increased use.

One of the hypothesized explanations for productivity slowdown in US since the past decade is that existing economic indicators (including GDP) do not properly measure the contributions of the latest wave to technological innovation, particularly digital goods and services. While average annual labor productivity growth was 2.8% per year over 1995-2004, it shrunk to 1.3% per year over 2005-2015 (Syverson 2016). An optimistic interpretation is that recent productivity gains due to innovations in IT-related goods and services are not properly reflected in the current productivity measures (e.g. Brynjolfsson and McAfee 2014, Aepfel 2015, Hatzius 2015). However, recent literature (Byrne et. al. 2016, Syverson 2016) has emphasized that while productivity mismeasurement may be important in recent years, it was also likely important in the past, so its power to explain the productivity slowdown is limited.

While motivated in part by this puzzle, our research does not aim to contribute directly towards this debate. Instead we focus on the more fundamental issue that GDP, and thus productivity, is not a direct measure of well-being in the first place. Thus, whether or not GDP or productivity mismeasurement has grown, is a distinct, albeit related, question from how well-being is changing. The gap between production (as measured by GDP) and well-being has been an issue at least since GDP was invented and, as we illustrate below, it is arguably an even bigger issue in the current digital era.

Consider the case of the music industry. Consumers shifted from buying physical units such as CDs, cassettes and vinyl records to downloading or streaming songs digitally through platforms such as iTunes, Pandora and Spotify. Digital goods have zero marginal cost and are

hence priced much lower (often even at zero) than physical goods. Between 2004 and 2008 consumers listened to more music (units of music purchased increased from under 1 billion to over 1.5 billion without counting illegal downloads) but the recording industry's revenues declined by 40% (Brynjolfsson and Saunders 2009) and this trend has continued. Moreover, Waldfogel (2012) provides compelling evidence that the quality of music has likely increased since 1999. Therefore, although the contribution of music industry to GDP statistics is shrinking, consumers are better off than before; they are listening to more and better music.

The relationships among GDP, consumer surplus and well-being can be understood by looking at three illustrative cases. First, consider a situation that roughly describes many classic physical goods such as cars, consumer surplus is more or less proportional to firm revenue (Figure 2). Keeping the supply curve fixed, as the size of the market increases as more consumers enter the market, the demand curve simply shifts right. In this case, both consumer surplus and quantity sold increase roughly proportionately.⁵ The increase quantity sold shows up in GDP statistics as sales increase, and hence both GDP and consumer welfare move in the same direction. At a given price, doubling the number of cars, apples or books sold is likely to roughly double revenues, GDP and consumer surplus. A similar logic applies for many services like haircuts, meals served or windows washed.

[Insert Figure 2 here]

A second category is the case of purely digital goods such as email, messaging apps, Facebook and Google search which have essentially zero marginal cost and are typically offered to the consumers for free. In some cases, digital goods earn revenues from advertising but this is an intermediate good, and does not contribute to GDP. However, changes in advertising revenues

⁵ In the special case of horizontal supply curve and thus constant price, the effect is exactly proportional.

are generally not closely related to changes in consumer surplus (Spence and Owen 1977).

Hence, as the demand for these free goods increases, consumer surplus will also increase but this change in well-being is not well-reflected in GDP (Figure 3). GDP may be completely unchanged due to this shift even though consumers are better off.

[Insert Figure 3 here]

A third case illustrates the situation faced by a number of traditional goods and services that are transitioning into digital goods and services. A good example of such a transition good is an encyclopedia. Since 2000s, people have increasingly flocked to Wikipedia to get information about a wide variety of topics updated in real time by volunteers. In 2012, Encyclopedia Britannica, which had been one of the most popular encyclopedias, ceased printing books after 244 years (Pepitone 2012). Wikipedia has over 60 times as many articles as Britannica had, and its accuracy has been found to be on a par with Britannica (Giles, 2005). Far more people use Wikipedia than ever used Britannica. While the revenues from Britannica sales were counted in GDP statistics, Wikipedia has virtually no revenues and therefore doesn't contribute anything to GDP other than a few minimal costs for running servers and related activities and some voluntary contributions to cover these costs. Similarly, many people now have digital maps, texting and other services available for no extra cost once they are able to access the Internet on mobile devices or home computers. For such transition goods, consumer surplus increases and revenue decreases as prices become zero (Figure 4). Hence GDP and consumer welfare move in opposite directions.

[Insert Figure 4 here]

More formally,⁶ consider the case of smartphones. Varian (2016) notes that a smartphone is a substitute (to varying degrees) for a camera, GPS, landline, gaming console, ebook reader, personal computer, video and audio player, maps/ atlas, alarm clock, calculator, sound recorder etc. Consider the simplifying case of two goods available in two periods: a digital camera and a feature phone in period 1 and a digital camera and a smartphone in period 2. Suppose that the value of the camera to the consumer is v_1 , the value of the simple feature phone is v_2 and the value of the smartphone is v_1+v_2 . Assume that a device fully depreciates in a time period, i.e. a consumer has to purchase new devices each period. Also assume that a consumer buys both the camera and the feature phone in period 1 and only the smartphone in period 2 and there are a total of x such consumers. Suppose that the price of the camera is p_1 in period 1, the price of the feature phone is p_2 in period 1 and the price of the smartphone is also p_2 in period 2. Therefore, we have

$$(1) (v_1-p_1)x + (v_2-p_2)x \geq 0$$

$$(2) (v_1+v_2-p_2)x \geq 0$$

Subtracting (1) from (2) gives us p_1x which is the change in total consumer surplus from period 1 to period 2. This simply corresponds to the cost savings of not buying the digital camera since it is now included in the smartphone. However the contribution of these goods towards GDP (i.e. the market price of final goods) is $(p_1+p_2)x$ in period 1 and p_2x in period 2. Hence change in GDP from period 1 to period 2 is $-p_1x$ which is exactly the opposite of change in consumer surplus. Therefore while GDP goes down due to people not purchasing the digital camera, consumer surplus goes up.

⁶ We thank Hal Varian for sharing his notes on GDP and Welfare, which contained this example.

In more concrete figures, the total number of digital cameras shipped worldwide dropped from 121 million units in 2010 to 24 million units in 2016.⁷ In the same period, worldwide smartphone sales increased from 297 million in 2010⁸ to 1.5 billion in 2016.⁹ Meanwhile, the total number of photos taken has also increased dramatically growing from 350 billion in 2010¹⁰ to 2.5 trillion in 2016.¹¹ Moreover, the price for taking a photo with your smartphone is essentially 0 (compared to a positive price for printing photos in the analog era).

Figures 3 and 4, along with the illustrative example of smartphone, suggest that changes in consumer surplus is an important supplement to GDP as a measure of well-being for the current digital economy for either transition goods or purely digital goods. This is likely to become increasingly relevant as more and more goods transition from physical to digital in a variety of areas including financial advising (robo advisors such as Wealthfront), customer service (AI powered services such as DigitalGenius) and law (AI powered bots such as DoNotPay).

While producer surplus cannot be inferred from consumer surplus, when it comes to technological advances, firms have typically been able to appropriate only a small fraction of the social returns (Nordhaus 2005). Accordingly, we can focus on consumer surplus. If the share of producer surplus contribution to the total social surplus remains relatively stable, then our results would have to be scaled up only slightly if one wanted to estimate total surplus. However, Furman and Orszag (2015) provide evidence that the top performing companies have been earning increasingly larger returns to capital. Therefore, measuring simply changes in consumer

⁷ http://www.cipa.jp/stats/dc_e.html

⁸ <http://www.gartner.com/newsroom/id/1543014>

⁹ <http://www.gartner.com/newsroom/id/3609817>

¹⁰ <https://www.nytimes.com/2015/07/23/arts/international/photos-photos-everywhere.html>

¹¹ <https://www2.deloitte.com/global/en/pages/technology-media-and-telecommunications/articles/tmt-pred16-telecomm-photo-sharing-trillions-and-rising.html>

surplus might underestimate changes in total surplus more significantly if the producer surplus grows relative to the consumer surplus.

2.2 Prior work measuring consumer surplus from digital goods

Recently there has been growing interest from researchers to estimate the changes in consumer surplus from digital goods. For instance, Greenstein and McDevitt (2011) estimate the additional consumer surplus created by broadband internet when consumers switched from dial-up to broadband. They estimate it to be between \$4.8 and \$6.7 billion from 1999-2006. For 2015, this figure is estimated to be \$55 billion (Syverson 2016). Although this approach captures the welfare gains due to better internet access, it does not capture the increasing value of the digital information goods available online.

Another stream of literature has tried to measure the value of digital information goods by measuring the time spent using them. The underlying assumption behind these papers is that there is an opportunity cost associated with using free digital goods and this cost is equal to the wages lost due to not working. Therefore, the value of these digital goods is equal to these lost wages. Using this approach, Goolsbee and Klenow (2006) estimate the effect of consumer gains from the internet for the median US resident to be \$3000 per year till the year 2005. Brynjolfsson and Oh (2012) extend this method to include substitutability between online and offline goods (e.g. TV). After accounting for this, they estimate the average annual change in consumer surplus of the internet to be about \$25 billion between 2007 and 2011.

Nakamura and Soloveichik (2015) estimate the value of free media by computing the online advertising revenues generated by websites. Including free media increases real GDP growth by 0.019% according to their estimates. However, advertising revenues do not capture

the entire value of the free digital goods. For example, in 2011 Google earned around \$36 billion ad revenue (Miller 2012) while Varian (2011) estimated the consumer surplus of Google to be between \$65-\$150 billion. Moreover, advertising revenues are not proportional to consumer surplus. Spence and Owen (1977) argue that advertisers pay for number of views regardless of whether these views created low or high value for a consumer. For example, advertising revenues can be high for a program of broad interest (more views) but welfare need not be very high because consumers might only be marginally interested in that program. Conversely, for a niche program which is valued very highly by a small group of consumers, welfare will be high but advertising revenues will be low.

While these estimates of consumer surplus are based on available market data, our method will use choice experiments to elicit consumers' valuation of goods. Specifically, we will ask consumers to make a choice between keeping a digital good or taking a monetary compensation when foregoing it. This approach experimentally varies the monetary values and therefore addresses the limitation that the actual market price of many digital goods is zero so that demand does not represent their value. Moreover, an experimental setting may be better able to isolate consumers' valuation of goods compared to market data that is typically confounded by many other variables; albeit, depending on the design of the experiment, it may come at the expense of being "hypothetical", i.e., inconsequential (Carson and Groves 2007) and therefore either noisy or biased, as we discuss below.

3. Methodology

3.1 Approaches to measuring consumer value

There are two general approaches to obtain input data to measure consumer value: 1) based on market data (“revealed preferences”) and 2) based on choice experiments or survey techniques (“stated preferences”).

Approaches based on revealed preferences analyze variation in observed market prices for a good to derive demand curves and price elasticity (e.g., Cohen et al., 2016; Greenwood and Kopecky 2011). Similarly, hedonic pricing models try to decompose the overall value of a good into the value contribution of its characteristics by applying regression-type models to the observed market prices and differences in characteristics of the goods (Williams 2008). However, both of these approaches require variance in the observed market prices and are therefore not directly applicable to goods that are provided for free. Alternatively, a variation on revealed preference can be provided when there is a proxy for market price, e.g., time spent using the digital goods (Goolsbee and Klenow 2006; Brynjolfsson and Oh 2012).

Stated preference elicitation techniques provide more flexibility because they do not require a market price or transactions to exist and can be applied to contingent scenarios (leading to contingent valuation studies). One approach to determining stated preference is to ask consumers directly about their maximum willingness-to-pay (WTP) in monetary terms. This question reveals a (potentially ratio-scaled) measure of a consumer’s value of the good. However, this type of question has been shown to be less reliable and less valid, likely because consumers are not used to formulating own prices and because they may feel an incentive to hide their true preferences (Miller et al. 2011; Carson and Groves 2007).

The introduction of non-hypothetical, incentive compatible variants to elicit WTP in form of auctions (e.g., Vickrey auctions, Vickrey, 1961) or lotteries (e.g., BDM, Becker, DeGroot, and Marschak 1964; Wertenbroch and Skiera 2002) mitigates some of these disadvantages, but at the expense of being more complex and by introducing (artificial) competitive pressure in auctions (Carson, Groves, and List 2014; Völckner 2006). These incentive compatible direct question formats may thus be ill-suited to either digital goods, in which supply is not restricted, or to large scale online choice experiments that consumers need to understand and answer quickly.

An alternative, indirect form of measuring stated preferences are discrete choice experiments (DCE) (Louviere, Hensher, and Swait 2000). DCEs ask consumers to choose between options and select the alternative that they value most. By experimental variation of the characteristics of the options (including prices) and applying logit or probit estimation models it is then possible to estimate consumers' utility function for the characteristics, i.e., their valuation of features and sensitivity to price changes. DCEs have become a common synonym for choice-based conjoint experiments that typically involve about 8 to 12 sequential choice tasks that present multiple alternatives, e.g., two to five, and each alternative varies in multiple attributes (Rao 2014). These DCEs have a long tradition in, among others, marketing (e.g., value of product features), transportation (e.g., valuation of travel time savings), contingent valuation (Carson et al. 2003), and are also applied to economic valuation contexts (e.g., Rosston, Savage, Waldman 2011). They are widely relied upon in the legal proceedings to estimate values of goods for the purposes of damages calculations (e.g. in the 2011-2014 Apple-Samsung lawsuit; see also McFadden 2014).

3.2 Proposed approach

We propose to measure consumer value of digital goods with DCEs. Instead of a conjoint-type experiment, we suggest a simpler implementation in which we only ask consumers to make a single choice among two options: Whether to keep access to a certain good or to give up the good and get paid a specific amount of money in return. We only ask one question per consumer and vary different price points systematically between consumers. The procedure can therefore be termed single binary discrete choice (SBDC) experiment (Carson and Groves 2007; Carson, Groves, and List 2014). We deliberately elicit only limited information from each consumer, i.e., data that is nominal-scaled, with the benefit that this information can be captured faster and more reliably. Consumers only have to make a decision between two options instead of thinking about a monetary figure themselves. Moreover, we can compensate for the loss in information at the individual level by using large-scale choice experiments and aggregating the responses from the overall sample in order to derive ratio-scaled demand data. Thus we use large (thousands of respondents), and potentially massive (hundreds of thousands or millions of respondents), sample sizes to overcome some of the limitations of earlier research relying on smaller samples.

3.3 Utility theory and choice model

DCEs in general, including SBDC questions, are compatible with economic theory and can be used to estimate neoclassical Hicksian welfare measures (McFadden 1974, Carson and Czajkowski 2014). We will use utility theory and the random utility model to conceptualize the value that individual consumers obtain from consuming digital goods and the monetary value that they attach to them.

Specifically, we represent the utility that a consumer experiences from consuming a digital good g by $U(g)$. In our SBDC questions, utility is only affected by a change in the availability of the good with consumption quantities restricted to 1 and 0, i.e., a consumer can either use a good within a defined time period (g^1) or not (g^0). We abstract away from the intensity or duration of usage in this conceptual model but can account for it in our empirical application. We assume a constant market price of zero for the goods, which therefore does not have to be added to the utility function. We also do not explicitly consider the influence of other attributes such as negative utility effects of advertising or limited privacy as they are nested within g^1 . These components can be easily added to the utility function when they are subject to experimental variation. We further assume that $U(g^1) \geq U(g^0)$, i.e., that consumers derive a non-negative utility of consuming the good (and would otherwise not use it). A measure of monetary value can then be estimated by introducing two Hicksian measures, either the compensating measure, C , or the equivalent measure, E , that have an effect on the consumer's income y (Carson and Czajkowski 2014), such that:

$$(3) U(g^1, y - C^*) = U(g^0, y), \text{ or}$$

$$(4) U(g^1, y) = U(g^0, y + E^*),$$

with $C > 0$ and $E > 0$.

C^* is typically referred to as willingness-to-pay (WTP) for getting access to the good, while E^* can be seen as willingness-to-accept (WTA) to forego it.

While, theoretically, C^* should have the same magnitude as E^* , empirical studies show that typically $E^* > C^*$, e.g., due to an endowment effect (Hanemann 1991; Kahneman, Knetsch, and Thaler 1990; Kahneman, Knetsch, and Thaler 1991). It therefore becomes relevant to define the *status quo* of the valuation approach. When valuing the availability of free digital goods it

seems reasonable to focus on WTA and assume that $U(g^l, y)$ is the *status quo* since using the good requires no upfront investment ($y - C$) from consumers.

When observing in the SBDC experiment that a consumer chooses to forego using a good for amount E instead of keeping it then we can assume that $U(g^0, y + E) > U(g^l, y)$, or $U(g^0, y + E) - U(g^l, y) > 0$. Therefore, only differential effects need to be considered between the choice options so that the overall income can be excluded and only the marginal effect of E needs to be considered. Without loss of generality, we can define the status quo utility as $U(g^l) = 0$. Consequently, a consumer will forego the good for amount E if $U(g^0, E)$ is positive, and will not if it is negative.

In order to estimate the equivalent measure E^* we need estimates of how valuable consumers find using the good and how sensitive they react to changes in E . The random utility model is the standard framework to estimate the underlying utilities. It assumes that utility U consists of a systematic component V and a random component e that is inherent to consumer choice behavior and/or unobservable to the researcher (Manski 1977; Thurstone 1927), such that $U(g^0, E) = V(g^0, E) + e$. Typically, it is assumed that the systematic utility consists of part-worth utilities for each of the goods components, i.e., $V = b_0 g^0 + b_1 E$. The framework then allows to express the observed choices as probabilities P within a binary logit model, i.e., the probability that a consumer chooses to forego the service (or, on an aggregate level, the share of consumers who are willing to accept E) is:

$$(5) P(g_0, E) = \exp(b_0 g^0 + b_1 E) / (1 + \exp(b_0 g^0 + b_1 E))$$

or $1 - P(g_0, E)$, for keeping the service. The parameters can be estimated using closed-form maximum likelihood procedures. The median equivalent measure E^* is then the price that makes

consumers indifferent between the two options so that $P(g_0, E^*) = 0.5$ or $b_0 g^0 + b_1 E = 0$, which leads to $E^* = -b_0 g^0 / b_1$.

Here, we represent the utility function as linear in terms of monetary amounts. We will relax this assumption in the empirical application to handle non-linear terms and include further demographic variables.

3.4 Criticism

Such SBDC questions have several advantages compared to approaches that directly ask consumers about their WTP or WTA (i.e., C or E). SBDC valuations are based on consumer choices that are most similar to day-to-day (purchase or consumption) activities. They are natural manifestations of consumers' preferences and are easy to accomplish. DCEs have been shown to achieve good (external) predictive validity and produce valid estimates of WTP so that they should be favored over direct elicitation techniques (Carson and Groves 2007; Miller et al. 2011, Wlömert and Eggers 2016).

SBDC questions are in line with economic theory and, based on the random utility model, can be used to estimate neoclassical Hicksian welfare measures (McFadden 1974, Carson and Czajkowski 2014). Moreover, a single “take-it-or-leave-it” referendum-like question has favorable incentive-compatibility properties, compared to multiple (e.g., “double-bounded”), sequential questions (Carson et al. 2003). Relatedly, single questions prevent the so-called starting point bias in follow-up questions, i.e., an anchoring effect in which subsequent prices are evaluated relatively to the first price the respondent was exposed to (Whitehead 2002).

However, the proposed SBDC questions, or rather contingent valuation questions in general, are not without criticism. Hausman (2012) identifies three major weaknesses of

contingent valuation questions: 1) differences between WTP and WTA, 2) hypothetical bias, and 3) inconsistencies regarding scope and embedding (see a detailed rebuttal to his criticism from Haab et al. 2013). While the choice experiments we run in this paper differ somewhat from the contingent valuation approach that Hausman discusses, they have enough similarities that it is worth considering his critiques in some detail.

Empirically WTP and WTA often do not give the same value, which is recognized as being inconsistent with neoclassical economic theory. Therefore, attempts have been made to extend (behavioral) theory in order to explain the disparities, e.g., with endowment effects, loss aversion, or uncertainty about the quality of the goods (Hanemann 1991; Kahneman et. al. 1991; Plott and Zeiler 2005).¹² As we argue above, we find that WTA better represents consumer welfare since free digital goods require no upfront investment from consumers. Moreover, research from behavioral economics suggests that measuring WTP instead of WTA for free goods can lead to biased estimates since consumers may take the market price of zero as an informational “anchor” so that WTP estimates are also biased towards zero (Ariely, Loewenstein, and Prelec 2003).

A hypothetical bias arises from SBDC questions (and contingent valuation questions) if consumers do not believe that their answer given to the stated preference question are consequential, e.g., that they actually need to forego any of the services in the near future and receive a compensation for it. Hence, consumers have no incentive to reveal their true valuation so that a random response would be as good as the true answer (Carson, Groves, and List 2014). According to the random utility model these answers increase the error variance so that

¹² Hausman (2012) argues that the gap between WTP and WTA is “likely due to the reality that answers to contingent valuation surveys do not actually reflect stable or well-defined preferences but instead are opinions invented on the fly” (p. 47). This statement, however, is at odds with the fact that the differences between WTP and WTA are persistent and consistent, which rules out random effects of unstable preferences but rather results from systematic differences in these question formats.

preference estimates are less reliable. If these errors are unbiased around the true value, then they will tend to cancel out as sample size grows. However, apart from random error, it has been shown that the hypothetical bias often leads to a *systematic* bias, e.g., due to strategic incentives to understate or overstate the true WTP (Carson and Groves 2007). This is a more serious concern, which, however, can potentially be addressed when the systematic processes are better understood.

In order to quantify the magnitude of the hypothetical bias we test it empirically by providing consumers with real money if they stop using Facebook. In that way, choices are consequential and consumers have a clear incentive to provide their true valuation of the service (Carson, Groves, and List 2014). Similar applications of SBDC questions date back to a study by Bishop and Heberlein (1979) in which they measure the consumer surplus of goose hunting permits by providing hunters with actual money for returning their hunting permits. Similar approaches to mitigate the hypothetical bias include the incentive alignment procedure applicable to DCEs that has been shown to increase (external) validity substantially (Ding 2007; Ding, Grewal, Liechty 2005; Wlömert and Eggers 2016). As discussed in our results section, we confirm that a hypothetical bias exists and that this criticism is justified. However, when looking at annual changes in consumer surplus differences between hypothetical and consequential approaches are less substantial.

The third major criticism, scope and embedding, refers to the proposition that consumers should be willing to pay more for a large effect than for a subset of that effect (or a good that is embedded in a larger package). Although these effects can be found empirically they are sometimes not considered large enough to be credible (Hausman 2012). Diamond and Hausman (1994) propose an adding-up test for the scope test. However, because digital goods can serve as

substitutes or complements (e.g., social media can provide messaging functions or video can be used to listen to music) the adding-up test is not appropriate in our context. Differences in scope and embedding therefore need to be analyzed in terms of the substitutability or complementarity of the goods, which we address in our empirical study.

Much of the criticism refers to the nature of non-market goods that are typically subject of contingent valuation studies (e.g., clean air or clean water). With these goods, which may be used predominantly in a passive way, consumers have limited to no active experience or are rather unfamiliar with their true value (Carson 2012; Carson and Czajkowski 2014; Hausman 2012). Digital goods, on the other hand, are typically used on a day-to-day basis and required an active step in getting access to them (e.g., the deliberate choice to subscribe to Facebook) so that consumers should be familiar with them and are therefore better able to express and quantify their value. Morwitz et al. (2007) support this notion in their meta-analysis and empirical study. We therefore expect that our research context mitigates much of the above-mentioned criticism. However, arguably, it does not eliminate them completely, which we will analyze empirically.

4. Consumer Surplus of Facebook

We use Facebook as a useful case in order to measure the consumer surplus with SBDC choice experiments. We benchmark the approach against a BDM lottery and explore its robustness in sensitivity analyses. In section 5, we apply the proposed SBDC approach to a broader list of goods and present an additional benchmarking study using best-worst scaling.

4.1 Incentive-compatible Single Binary Discrete Choice Experiment

In order to circumvent a hypothetical bias we applied the SBDC experiment in a non-hypothetical, incentive compatible procedure to measure the consumer surplus of Facebook. We asked consumers if they would prefer to 1) keep access to the Facebook or 2) give up Facebook for one month¹³ and get paid $\$E$. We varied $\$E$ across twelve¹⁴ price points ($E = 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 1000$). To make the SBDC question consequential for the consumer, we informed them that we will randomly pick one consumer out of every 200 respondents¹⁵ and fulfill that person's selection. Specifically, we told respondents that if they choose "Keep access to Facebook" nothing will change for them, however, they will also not receive any money. If they choose "Give up Facebook and get paid $\$E$ ", we promised them the money in cash provided that they do not access Facebook for one month. We further informed them about the procedure how we can monitor their Facebook online status remotely and the requirement to provide their email address (see Figure A.1 in the appendix for the exact question wording and monitoring process).

We recruited consumers for this study from Peanut Labs, a professional panel provider with 2.9 million active panelists and member of several survey research organizations, including CASRO, ESOMAR, and MRA (Peanut Labs 2015). We invited respondents in June/July 2016 and 2017 to be able to measure annual changes. We targeted consumers that were 18 years or older and lived in the US. We further asked consumers to select all online services they have

¹³ We restricted the time frame to one month in order to keep the incentive compatibility procedure manageable. We address the sensitivity of the valuation depending on the time frame in the sensitivity analysis.

¹⁴ In a follow-up study, we included additional price points, i.e., \$0.01, \$5, \$200, \$500 and found consistent results.

¹⁵ Carson, Groves, and List (2014) show that stochastically binding procedures (here: one out of every 200 respondents) do not significantly affect the results compared to deterministically binding procedures. We can confirm this result for our Facebook study in which we also tested a condition in which one out of every 50 respondents was selected (E was kept at \$50 in this condition). We did not find significant differences in the choice behavior when varying the chances to win ($p = 0.236$).

used in the last twelve months from a list of 14 options, including a non-existent online service. Consumers had to select Facebook in order to qualify for the survey; if they (also) selected the nonexistent service which we included in the survey they were disqualified. We have set quotas for gender, age, and US regions to match US census data (File and Ryan 2014) and applied post-stratification for education and household income.

Consumers who accessed the survey were randomly allocated to one of the tested price points. We sampled the highest and lowest price points twice as often in order to obtain more reliable estimates for the endpoints of the demand function. We received 2885 complete responses ($n_{2016} = 1497$, $n_{2017} = 1388$).

Figure 5 plots the estimated WTA demand curves, separated for 2016 and 2017.¹⁶

[Insert Figure 5 here]

In order to measure the WTA and quantify the annual change we estimated a binary logit model that accounts for the magnitude of E (here, $\log(E)$ provided a better fit to the data), year (dummy variable), and whether the samples in the different years differ in sensitivity towards E . Table 1 shows the estimation results. The intercept represents the share of consumers in 2016 who prefer to keep Facebook at $E = \$1$ (i.e., $\log(1) = 0$). This share is estimated to be $\exp(1.2)/(1 + \exp(1.2)) = 76.9\%$. This share is non-significantly larger in 2017 ($p = 0.166$) with $\exp(1.2 + 0.29)/(1 + \exp(1.2 + 0.29)) = 81.6\%$. In 2016, the sample's utility decreased by -0.309 with every one-unit increase in $\log(E)$, leading to a median $WTA_{2016} = \$48.49$ per month. This means that 50% of the Facebook users in our sample would give up all access to Facebook for one month if we paid them about \$50 or more. The Facebook users in 2017 reacted significantly more sensitive towards differences in E ($p = 0.049$). A one-unit increase in $\log(E)$ results in a utility

¹⁶ The plots show the shares of consumers who prefer to keep using Facebook instead of being willing to accept the money in order to be consistent to normal practice for representing demand curves. That is, we plotted them in a way that makes it easier to see the negative effect of price.

decrease of $-0.309 - 0.101 = -0.410$. As a consequence, consumers in 2017 were willing to accept a lower amount to give up Facebook, i.e., median $WTA_{2017} = \$37.76$ per month. Since the sample consists of Facebook users only a surplus measure also needs to consider the overall amount of consumers who use Facebook. However, the share of Facebook users in the US increased from 2016 to 2017 by just 2.6%¹⁷, which cannot offset the negative tendency in median WTA.

We used bootstrapping to calculate 95% confidence intervals for the median WTA values, i.e., $CI_{2016} = [\$32.04, \$72.24]$, $CI_{2017} = [\$27.19, \$51.97]$. The range of the confidence intervals illustrates the limitation of the approach in being less precise, given the current sample size. Although the median WTA values suggest a substantial drop in value the confidence intervals are very broad so we can't reasonably rule out that this is simply due to chance. We address the effect on precision by using larger sample sizes in the sensitivity analyses below.

[Insert Table 1 here]

We added usage and demographic variables to further understand differences in consumer value. The estimation results can be found in Table 2. The usage of Facebook per week (self-reported, measured on a 5-point scale from “less than 1 hour” to “more than 14 hours”) is a significant predictor for the value of Facebook ($p = 0.006$). The more time a consumer spends on Facebook the more likely they are to keep their access. Similarly, the more friends someone has on Facebook (self-reported, measured on a 6-point scale from “less than 50” to “more than 1000”) the more compensation they require to leave Facebook ($p = 0.024$). In terms of activities on Facebook (measured on a 6-point scale ranging from “never” to “several times a day,” consumers perceive significantly more value in Facebook the more they post status

¹⁷ <https://www.statista.com/statistics/408971/number-of-us-facebook-users/>

updates or share pictures and videos ($p = 0.010$), the more they like and comment ($p = 0.018$), and play games ($p = 0.025$). Watching videos is marginally significant ($p = 0.080$), while using the messenger and chat provides no additional value ($p = 0.100$). Consistently, we find significant substitution effects due other social media services, i.e., Instagram ($p = 0.025$), and video platforms, i.e., YouTube ($p = 0.003$). Thus, consumers who also use Instagram or YouTube are more likely to give up Facebook. Services that are not related to activities that provide value on Facebook show no significant substitution effects (e.g., Wikipedia, $p = 0.601$).

In terms of socio-demographics, we find significant effects for gender and age of the respondent, as well as household income. Specifically, we see that female respondents are more likely to keep Facebook than male users ($p = 0.011$). The same holds for older consumers ($p < 0.001$). The effects for household income are less consistent. Households with an income between 100K and 150K perceive significantly less value in Facebook ($p = 0.019$), while higher income households value Facebook more ($p = 0.008$). The effect is also significantly positive for consumers who preferred not to disclose their income ($p = 0.004$). Education and US region are not significant (not shown in Table 2).

[Insert Table 2 here]

To summarize, the SBDC experiment leads to plausible demand functions and plausible effects of usage and demographic variables. The results indicate that Facebook provides substantial value to consumers who would require a median compensation of about \$50 per month for leaving this service. We find no evidence that this valuation increased from 2016 to 2017; if anything it appears to have declined somewhat. However, given the nature of choice data the estimated median WTA values are limited in terms of precision compared to directly elicited values, which we will use as a benchmark method in the next section.

4.2 Benchmark method: BDM lottery

As a benchmark to check the convergent validity of the SBDC approach, we applied an incentive compatible BDM lottery procedure (Becker, DeGroot, and Marschak 1964) in order to elicit direct, numeric responses from consumers about their WTA. Specifically, we asked consumers about the minimum amount of money they would request in order to give up Facebook for one month. In order to achieve incentive compatibility we informed respondents that the amount will serve as their bid in a lottery. The BDM lottery process instructs that, after the survey, a random price will be drawn from a uniform distribution of values. If the random price is higher than the bid, the respondent will be paid the random price when giving up Facebook for one month. If the random price is lower than the bid, the respondent will receive no money but can keep the access to Facebook. Thus, the rational, utility-maximizing strategy for the respondent is to bid exactly their true value for Facebook.

We conducted the BDM lottery in the lab of a European university, parallel to an incentive compatible SBDC experiment. The lab setting allowed us to explain the BDM procedure in detail and make sure that the respondents understood the mechanism. In total, 139 students took part in the lottery. We compare this sample to a sample of respondents that took part in the incentive compatible SBDC experiment of the lab ($n = 356$). The SBDC procedure was identical to the Peanut Labs study but used monetary offers in €. Figure 6 shows the estimated demand functions that result from both approaches. The SBDC derived function is closely aligned to the BDM demand function. The observed shares correlate strongly (Correl. = 0.891). Fitting a regression model to the observed shares ($R^2 = 0.755$) shows that the BDM approach estimates a larger intercept than the SBDC approach ($p = 0.013$), i.e., more respondents are willing to keep Facebook at low monetary values. This is plausible since BDM allows

respondents to have more control over their bids and few respondents expressed to accept low monetary values, while the SBDC approach follows a take-it-or-leave-it mechanism with exogenous monetary offers. More importantly, however, both approaches do not differ in the estimated price sensitivity ($p = 0.278$). While this result gives us confidence in our estimates from the SBDC experiment, we explore its robustness in further sensitivity analyses.

[Insert Figure 6 here]

4.3 Sensitivity analyses

We assess the robustness of the SBDC approach regarding its sensitivity to a hypothetical bias, random responses, sample size, and the analyzed time frame.

4.3.1 Hypothetical bias

In order to measure the hypothetical bias we applied a hypothetical scenario parallel to the incentive compatible SBDC experiments in section 4.1. Specifically, we conducted the same surveys as in the incentive compatible scenarios with Peanut Labs in June/July 2016 and 2017 but without informing consumers that their answers were consequential. We allocated respondents randomly to the incentive compatible (IC) and non-incentive-compatible (NIC) scenarios. In addition to the 2885 respondents in the IC studies, we interviewed 2878 consumers in the NIC conditions ($n_{2016,NIC} = 1500$, $n_{2017,NIC} = 1378$).

For illustration, we detail the results for the 2016 study first. Figure 7 compares the observed shares between IC and NIC groups. For very low prices, i.e., a price of \$1, the IC and NIC condition produce almost identical shares, which is reasonable. For higher prices the disparities increase leading to consistently higher shares in the IC condition. The estimation of

the binary logit model confirms that the IC consumers do not differ in the intercept ($p = 0.905$) but they react significantly less sensitive towards differences in E ($p = 0.002$, see Table 3). Consequently, the IC consumers are less attracted by the monetary offers and require a significantly higher amount in order to give up Facebook ($WTA_{IC,2016} = \$48.49$, $CI_{IC,2016} = [\$32.04, \$72.24]$). Consumers in the NIC setting are satisfied with lower amounts, i.e., $WTA_{NIC,2016} = \$13.80$ per month (95% $CI_{NIC,2016} = [\$9.80, \$19.19]$). Consequently, the hypothetical WTA is understated in this research context and needs to be calibrated by a factor of 3.5.

[Insert Figure 7 here]

[Insert Table 3 here]

The results for the 2017 study are consistent. In this case, the median WTA in the NIC condition is $\$9.18$ (95% $CI_{NIC,2017} = [\$6.07, \$13.70]$), compared to $\$37.76$ in the IC scenario ($CI_{IC,2017} = [\$27.19, \$51.97]$), which leads to a calibration factor of 4.1 (see appendix, Table A.1 for the full estimation model that accounts for year and group membership).

Our results suggest that the hypothetical bias can be substantial. More importantly, however, our primary interest is not the absolute amount of consumer surplus for Facebook but annual changes in value. In this case, the incentive compatible study would estimate a loss in value of -20.1% from 2016 to 2017, while the hypothetical study calculates a loss of -31.7%. Despite the hypothetical bias, the annual changes move in the same direction and are more closely aligned than the absolute valuations.

4.3.3 Effect of random answers

Random answers increase the error variance in choice model estimations. The error variance, in turn, has a negative effect on the precision, i.e., scale of the estimates S in logit choice models (Hauser, Eggers, Selove 2016). Specifically, the scale S is inversely proportional to the error variance. The scale S cannot be separately identified, such that it is incorporated in the “raw” estimated utilities b :

$$V = (S * b_0) g^0 + (S * b_1) E.$$

Lower scaled estimates (more error), i.e., estimates with lower magnitude, cause the logit function to become more linear. Higher scaled estimates (less error) lead to a stepwise function that allows to predict decisions and identify the median WTA more precisely (see Figure 8).

[Insert Figure 8 here]

The effect is demonstrated empirically in Table 4. The table shows the result of a modified bootstrapping procedure in which 1,000 subsamples were drawn from the 2016 IC Facebook sample for illustration¹⁸. In each subsample we replaced R randomly selected original responses with the same amount of *random* answers and re-estimated the logit model. The results show that more random noise in the answers decreases the scale of the estimates. The scale S is proportional to the relative share of non-random answers. Having more random answers than original responses ($R = 800$) causes the magnitude of the estimates to be less than half the size of the original estimation without additional random answers ($R = 0$). However, the median WTA (averaged across the 1,000 subsamples) as well as the absolute standard error of the estimates remain largely unaffected. Surplus measures that consider the overall demand function by integrating the demand function, here in the interval from \$1 to \$1000, are biased by random answers. We therefore only report WTA measures in our analyses.

¹⁸ We obtain similar results for the NIC group and for the 2017 samples.

[Insert Table 4 here]

The simulated results illustrate that the magnitude of error is not a valid reason to explain the hypothetical bias. Interestingly, when we compare how many of the observed choices can be predicted correctly based on the estimated model, we find a better fit for the NIC group (hit rate_{NIC} = 69.9%) than for the IC group (hit rate_{IC} = 62.1%). This suggests that consumers in the IC group faced a decision that was more difficult to make, likely because their choices were consequential. It is important to note that the misclassified choices are not necessarily due to purely random responses. These cases can also be explained by heterogeneity among consumers that is not accounted for in the estimation models, either with respect to their valuation of Facebook or regarding their general price sensitivity (or both).

We find evidence that few respondents answer purely at random. Using GCS (n = 502), we asked a question in which we requested respondents to select all services that they have used in the last 12 months. Only 1% of the respondents have chosen a non-existing service, i.e., answered randomly.

4.3.4 Effect of sample size

Next to random noise, the precision of the WTA estimates also depends on the sample size. To analyze the magnitude of the effect we used bootstrapping with varying subsample sizes to observe the effect on standard errors and confidence intervals for the WTA estimate. Each subsample of a given size was again randomly drawn 1000 times from the original sample (IC group in 2016). As expected, Table 5 demonstrates that the standard errors of the estimates are reduced by the square-root of 2 when doubling the sample size (in this case the scale of the estimates remains largely unaffected). This general pattern also holds for the standard error of

the WTA estimate. However, since WTA is a ratio of two stochastic variables this generalization is approximate. The results show how the 95% confidence interval narrows when increasing the sample size. There is uncertainty in the measure even with a sample size of 1500. A 95% confidence interval of $\pm\$10$ would be achieved with a sample of 6000 consumers. This result highlights the need for large-scale, and potentially “massive”, sample sizes to measure consumer surplus precisely.

[Insert Table 5 here]

4.3.2 Effect of the analyzed time frame

In the previous incentive compatible studies we used one month instead of one year as the time frame that respondents should forego Facebook. This raises the question to what extent consumers are sensitive to the time frame. To address this question, we conducted SBDC experiments in an incentive compatible setting in which, in addition to prices E , we varied the time frame across three periods, i.e., $T = 1$ week, 2 weeks, 1 month. We recruited another sample from Peanut Labs in 2017 using the same criteria as in the previous studies, however, we did not screen out respondents who do not use Facebook (assuming that these respondents would accept any low monetary compensation; empirical valuations are therefore lower than in the previous study). A total of 1499 respondents were available for the analysis.

Table 6 shows the estimation results. As expected, the time frame has a significant positive effect on the probability to keep Facebook. Accordingly, the median WTAs for the different time frames are \$3.92 for one week, \$10.53 for two weeks, and \$17.61 for one month. These values and the beta estimates suggest that the effect of time might not necessarily be linear.

[Insert Table 6 here]

In order to get a better overview of the effect of time we sampled 5021 additional respondents in a hypothetical setting using Google Consumer Surveys (see section 5). We allocated these respondents randomly to one of ten conditions that differ in the time frame: $T = 1$ hour, 1 day, 1 week, 2 weeks, 3 weeks, 4 weeks, 1 month, 2 months, 3 months, 6 months, 1 year (operationalized in the estimation model in terms of number of days). We kept E constant at \$50 in this study. Figure 9 shows the observed shares of respondents who prefer to keep Facebook at the different time frames and the predicted time function according to the binary logit model (using $\log(T)$ and $\log(T)^2$ as predictors, see Table A.2 in the appendix). It confirms a positive log-linear effect of time with increasing marginal effects. Accordingly, consumers are more likely to keep Facebook the longer the time frame and this effect is reinforced with increasing duration. We use a time frame of one year in the large-scale studies we present next.

[Insert Figure 9 here]

5. Large-scale Studies to Measure Consumer Surplus

5.1 Google Consumer Surveys: Single Binary Discrete Choices

For the implementation of our large-scale studies, we use Google Consumer Surveys (GCS) as our primary platform. GCS allows us to run short one-question surveys cheaply and quickly and is therefore well suited for our SBDC experiments. A number of online publishers (including news and arts/ entertainment sites) participate in GCS and host these choice experiments on their site as a gateway to access premium content (Stephens-Davidowitz and Varian 2015). Users have to answer the survey in order to unlock premium content (Figure 10).

Survey creators pay per response, part of which goes to the publisher for hosting it. In addition to the responses, some demographic characteristics of the respondents such as region, age, gender and income are also provided which are inferred from IP address, location, browsing history (provided by Google's DoubleClick cookies which are also used to serve ads) and census data. Prior research has found that GCS results are very similar to those obtained from other surveys conducted by professional organizations such as Pew (Stephens-Davidowitz and Varian 2015)¹⁹.

[Insert Figure 10 here]

We identified the most widely used apps and websites on various devices and combined them into the following eight categories: Email, Search Engines, Maps, E-commerce, Video, Music, Social Media, and Instant Messaging. We ran SBDC surveys for each of these categories in June/July 2016 and 2017. In these studies we asked consumers to consider giving up access to these categories for one year. As a compensation, we chose 6-15 price levels for each category and around 500 responses per price level per year. If the median WTA was outside the range of our initial set of price levels, we increased the number of price levels in the following year in order to accommodate higher prices (for Search Engines, Email, Maps).

The observed shares and estimated demand curves are shown in Figure 11. The demand curves appear plausible and are consistent across time (solid lines represent 2016, dashed lines 2017). The annual changes suggest an increase in the valuation for these categories, albeit being small. This notion is confirmed when inspecting the median annual WTA values per year in Table 7. As in the Facebook study, the range of the confidence intervals is large so that the significance of the changes cannot be estimated reliably.

¹⁹ To confirm that there is no selection bias we compared the NIC group from the Peanut Labs sample (see section 4.3.1) to a GCS sample ($n = 1451$). Because Google Surveys do not screen respondents if they are Facebook users or not, unlike in the Peanut Labs study, we matched the NIC group by accounting for the share of non-Facebook users. A binary logit model confirms that there are no significant differences between both samples, neither in terms of their intercept ($p = 0.991$) nor sensitivity towards E ($p = 0.474$). See appendix for details (Table A.3, Figure A.2).

According to the median WTA estimates for 2017, Search Engines (\$17,530) is the most valued category of digital goods followed by Email (\$8,414) and digital Maps (\$3,648). One possible reason that these values are high relative to the other goods in our analysis may be the lack of effective substitutes for search engines, email or digital maps compared to the other categories in our sample. Since most consumers do not directly pay for these services, almost all of the WTA for these goods contributes towards consumer surplus. What's more, for many people, these services are essential to their jobs, making them reluctant to give up these goods.

Video streaming services (e.g. Youtube, Netflix) are valued by consumers with a median WTA of \$1,173 per year. Some consumers do pay for some of these services. However, these amounts are of the order of \$10-\$20 per month, or \$120-\$240 per year (for those who pay). Our measure suggests that the surplus the median consumers receives from these goods is a 5-10 multiple of what they actually pay (and which can show up in national accounts). The remaining categories for which we estimated the median WTA are (in descending order) E-Commerce (\$842), Social Media (\$322), Music (\$168), and Instant Messaging (\$155).

All these estimates are potentially biased downwards due to lack of incentive compatibility in these studies. Nevertheless, the sum of these estimates suggests there is a significant amount of consumer surplus from digital goods and a positive tendency over time.

[Insert Figure 11 here]

[Insert Table 7 here]

The available demographic variables (gender, age, income and urban density) reported by Google were added to an extended model to determine effects for different consumer segments. These extended logit models are reported in Table A.4 in the appendix. This reveals a number of patterns that are interesting and may have implications and for research and business.

For instance: The value of search engines increases by age and income and is higher for female consumers. Similar effects of age and gender can be observed for the email category. In this case, consumers in urban areas and with a median income of \$50K to \$75K also perceive a higher value. For maps the effect of age on WTA follows an inverse U-shape. Middle-aged consumers of 35-44 years value maps most. Income has a positive effect on the valuation of maps. A similar inverse U-shaped effect between age and valuation can be seen for e-commerce. In this case, the maximum value is experienced by 55-64 year-old consumers. In addition, female consumers perceive a higher value from online shopping. Age has a negative effect for the video and music categories. While this trend is consistent across all age groups for videos, the negative trend only starts at an age of 45 years or older for music. Male consumers value videos more. The music category is preferred in urban areas. In the social media category only gender shows a significant effect such that female users value this category more. The same holds for instant messaging. In this category, the youngest age group (18-24 years) perceives the highest value. Older consumers perceive significantly less value. Our approach opens the door to testing a variety of hypotheses and uncovering most such patterns relatively easily.

Our approach can be used for digital and non-digital goods alike. As an example, we also ran SBDC surveys to estimate the WTA to give up the option of eating breakfast cereal²⁰ for one year. Figure 12 plots the WTA demand curve for breakfast cereal. We estimate the median WTA to give up breakfast cereal to be \$44.27 in the US in 2017 (95% CI₂₀₁₇: [\$37.19; \$52.47]).²¹ This estimate is almost identical to the results from 2016 (95% CI₂₀₁₆: [\$37.98; \$49.74]). Examining non-digital goods can help us calibrate the relative importance of some of the digital goods we

²⁰ Economists have studied this industry using a variety of approaches. See e.g. Hausman, 1996, Schmalensee, 1978, Nevo, 2001, and others. Hausman (1996) estimates the consumer surplus due to entry of a new cereal brand (Apple-Cinnamon Cheerios) to be \$0.3136 per person per year.

²¹ This figure is in addition to the price paid by consumers for buying breakfast cereal.

examine. We therefore also incorporate non-digital goods in the benchmark study using best-worst scaling.

[Insert Figure 12 here]

5.2 Benchmark method: Best-worst scaling

As a benchmark to GCS we conducted additional choice experiments based on the best-worst scaling approach (Flynn et al. 2007; Marley and Louviere 2005). Best-worst scaling asks consumers to repeatedly select the best and worst options from sets of alternatives. Collecting more information, both within the choice set and across sequential choice sets, for each consumer makes this approach more efficient compared to the SBDC approach, which elicits only one decision. Moreover, consumers are required to make a tradeoff when deciding which goods they perceive as most and least valuable. This may mitigate or even eliminate the systematic hypothetical bias, at least with respect to the ordinal ranking of the choices.

We used a list of 34 goods (a mixture of digital and non-digital goods), including nine price points ranging from \$1 to \$20,000, that consumers had to evaluate. Since we examined the value of not having access to specific services or amenities for one year the prices were also formulated as losses in order to be comparable, e.g., “earning \$10,000 less for 1 year.” The price sensitivity we are observing is therefore closer to WTP than WTA.

We used three options within each choice set for each individual so that respondents created a full ranking of the three options in a set by indicating the best and worst options. Figure 13 shows an example of such a choice set. Respondents answered 10 or 11 sets²² in order to be

²² We used two subsamples that differed in the number of goods and number of choice sets in order to accommodate different price points. One subsample (n = 204) evaluated 30 options in 10 choice sets; the other subsample (n = 299) 33 options in 11 sets.

exposed to each good. We randomized the allocation of goods and prices to choice sets across respondents.

[Insert Figure 13 here]

We recruited consumers for this online study via Peanut Labs in 2017. We targeted consumers that were 18 years or older and lived in the US. Consumers who did not fulfill these criteria were screened out. We controlled quotas for gender, age, and US regions to match US census data (File and Ryan 2014). In total, 503 respondents completed the study.

We estimated utility parameters using a multinomial logit model. We considered both best and worst choices in the same model by interpreting utilities from best choices as the negative of worst choices. The estimation leads to interval-scaled utility scores that represent the disutility of not having access to the 34 goods (or earning less income) for one year, which are depicted in Figure 14 (see also Table A.5 in the appendix). We have set the lowest ranked service for the US, WhatsApp, as a reference category so that utilities are expressed relative to WhatsApp. The ranking of the goods is consistent to the SBDC experiments for the eight most widely used categories using GCS, with only one exception: online shopping is valued more than maps and video streaming in the best-worst scaling approach, while we find it to be valued less in the GCS surveys. When comparing the utilities of the services to the utility scores of the price levels we find, as expected, consistently lower implied WTP values than WTA estimates according to the GCS survey. Estimating a demand function and interpolating WTP shows very strong correlation among BWS and SBDC valuations (Correl. = 0.911). Overall, comparing the results of both approaches indicates convergent validity.

[Insert Figure 14 here]

6. Discussion

With advances in information technologies, we can now gather data at a large scale in close to real time. Initiatives such as MIT’s Billion Prices project²³ and Adobe’s Digital Price Index²⁴ are collecting price data from online retailers in real time to compute price and inflation indices. We explore the potential to reinvent and supplement the measurement of economic well-being by taking advantage of the ease of gathering data in the digital era. The end goal of this research agenda is to design a scalable method of measuring changes in consumer surplus due to technological advancements. We explore a potential way of measuring changes in consumer surplus through SBDC experiments. Our method is highly scalable and relatively inexpensive. Therefore, it can be run at very frequent, regular intervals to keep track of changes in consumer surplus. As argued previously, this measure can be an important complementary indicator of consumer well-being for the digital economy.

In a series of online experiments we show that the SBDC approach leads to plausible demand functions that are consistent with other validated approaches. We find that free digital goods provide substantial value to consumers even if they don’t contribute significantly to GDP. We further find that our approach can detect consumers’ sensitivity towards different time frames, e.g., whether consumers use (or not use) the goods for one week, one month, or one year. We find that time has a positive effect on the probability to keep a service with increasing marginal returns. Some consumers seem to be willing to undergo “digital detox” for a short duration by giving up internet or individual services like Facebook either through self-control or by installing software which blocks particular sites. This might explain consumers weaker

²³ <http://bpp.mit.edu/>

²⁴ <https://blogs.adobe.com/digitalmarketing/analytics/introducing-digital-economy-project/>

sensitivity towards short time frames and raises interesting questions about neoclassical economic models of rational choice, self-control and the nature of utility functions.²⁵ Due to this trend we recommend to use longer time frames for the evaluation, e.g., one year.

In order to address the limitation of a hypothetical bias of the proposed approach we have compared consumers' valuation of Facebook in an incentive compatible and hypothetical setting. We confirm that a hypothetical bias exists such that valuations of Facebook in the hypothetical scenarios tend to be significantly underestimated. The generalizability of such correction factors needs to be analyzed further in future studies. However, the differences between hypothetical and incentive compatible approaches are much less severe when analyzing annual changes in value.

A major limitation of our study remains the lack of precision in our estimates. While the BEA is able to measure GDP very precisely (e.g. US GDP was reported as \$16,514,593,000 on the first day of 2016), we are only able to provide a relatively coarse estimate of changes in consumer surplus, even in our large-scale studies. Future applications should use larger, i.e., *massive*, sample sizes to narrow the confidence interval of the WTA estimates or could explore adaptive approaches that adjust the analyzed price intervals dynamically in order to find the relevant price range for the median WTA.

While the median WTA is robust to random noise in the data, the overall demand functions are not. Reporting the median, however, would limit the application of the SBDC approach to goods that at least 50% of the consumers (or consumer segments) are using. Alternatively, future research could report other key percentiles, e.g. the valuation for people at the 90th percentile, or other benchmarks, when comparing goods to each other. Before being able

²⁵ Economics continues to evolve to take account mental biases that deviate from traditional notions of rationality, e.g. Kahneman et. al. 1990, Kahneman 2011, Thaler 2017.

to derive surplus measures along the overall demand curve (e.g. as in Greenwood and Kopecky 2011) we need further evidence to confirm that the error variance in the data remains consistent over time and therefore cancels out when calculating annual changes.

Another limitation of our study is that it is biased towards people using the internet. Our choice experiments are only accessible online, therefore people not using the internet at all are excluded. Pew estimates that about 15% of Americans don't use the internet.²⁶ Accordingly, our results must be interpreted as relevant to this audience, but not necessarily others.

That said, our approach is at least attempting to directly measure a concept that we know is not correctly measured by other official data. In short, we believe it is better to be approximately correct than precisely wrong.

²⁶ <http://www.pewresearch.org/fact-tank/2015/07/28/15-of-americans-dont-use-the-internet-who-are-they/>

References

- Ariely, Dan; Loewenstein, George and Prelec, Drazen. “ ‘Coherent Arbitrariness’: Stable Demand Curves without Stable Preferences.” *Quarterly Journal of Economics*, February 2003, 118 (1), pp. 73-105.
- Bates, W. (2009). Gross national happiness. *Asian-Pacific Economic Literature*, 23(2), 1-16.
- Becker, G.M., DeGroot, M.H., & Marschak, J. (1964). Measuring Utility by a Single-Response Sequential Method. *Behavioral Science*, 9, 226–232.
- Bishop, R.C. and Heberlein, T.A., 1979. Measuring values of extramarket goods: Are indirect measures biased?. *American journal of agricultural economics*, 61(5), pp.926-930.
- Brynjolfsson, E. and Oh, J., 2012. The attention economy: measuring the value of free digital services on the Internet. *Proceedings of the 33rd International Conference on Information Systems*, Orlando, December 2012.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.
- Brynjolfsson, E., & Saunders, A. (2009). What the GDP Gets Wrong.(Why Managers Should Care). *MIT Sloan Management Review*, 51(1), 95.
- Carson, R. and Czajkowski, M., 2014. The discrete choice experiment approach to environmental contingent valuation, in: Hess, S. and Daly, A. eds., 2014. *Handbook of choice modelling*. Edward Elgar Publishing.
- Carson, R.T., 2012. Contingent valuation: A practical alternative when prices aren't available. *The Journal of Economic Perspectives*, 26(4), pp.27-42.
- Carson, R.T., Groves, T., 2007. Incentive and informational properties of preference questions, *Environmental and Resource Economics*, 37(1), pp.181-210.

Carson, R.T., Groves, T., List, J.A., 2014. Consequentiality: A theoretical and experimental exploration of a single binary choice, *Journal of the Association of Environmental and Resource Economists*, 1(1/2), pp.171-207.

Carson, R.T., Mitchell, R.C., Hanemann, M., Kopp, R.J., Presser, S. and Ruud, P.A., 2003. Contingent valuation and lost passive use: damages from the Exxon Valdez oil spill. *Environmental and resource economics*, 25(3), pp.257-286.

Cavallo, A., & Rigobon, R. (2016). The Billion Prices Project: Using online prices for measurement and research. *The Journal of Economic Perspectives*, 30(2), 151-178.

Cohen, P., Hahn, R., Hall, J., Levitt, S., & Metcalfe, R. (2016). Using Big Data to Estimate Consumer Surplus: The Case of Uber (No. w22627). National Bureau of Economic Research.

Diamond, P.A. and Hausman, J.A., 1994. Contingent valuation: Is some number better than no number?. *The Journal of economic perspectives*, 8(4), pp.45-64.

Ding, M., 2007. An incentive-aligned mechanism for conjoint analysis. *Journal of Marketing Research*, 44(2), pp.214-223.

Ding, M., Grewal, R. and Liechty, J., 2005. Incentive-aligned conjoint analysis. *Journal of marketing research*, 42(1), pp.67-82.

File, T. and Ryan, C. 2014. Computer and Internet Use in the United States: 2013, U.S. CENSUS BUREAU. (Accessed at: <http://www.census.gov/history/pdf/2013computeruse.pdf>)

Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2007). Best–worst scaling: what it can do for health care research and how to do it. *Journal of Health Economics*, 26(1), 171-189.

Furman, J., & Orszag, P. (2015). A firm-level perspective on the role of rents in the rise in inequality. Presentation at “A Just Society” Centennial Event in Honor of Joseph Stiglitz Columbia University.

Giles, J. (2005). "Internet encyclopaedias go head to head: Jimmy Wales' Wikipedia comes close to Britannica in terms of the accuracy of its science entries". *Nature* 438 (7070): 900–1

Goolsbee, A., & Klenow, P. J. (2006). Valuing Consumer Products by the Time Spent Using Them: An Application to the Internet. *The American Economic Review*, 96(2), 108-113.

Greenstein, S. and McDevitt, R.C., 2011. The broadband bonus: Estimating broadband Internet's economic value. *Telecommunications Policy*, 35(7), pp.617-632.

Greenwood, J. and Kopecky, K.A. 2011. "Measuring the Welfare Gain from Personal Computers: A Macroeconomic Approach" Federal Reserve Bank of Atlanta Working Paper No. 2011-05. (Accessed at SSRN: <http://ssrn.com/abstract=1800266>)

Haab, T.C., Interis, M.G., Petrolia, D.R. and Whitehead, J.C., 2013. From hopeless to curious? Thoughts on Hausman's "dubious to hopeless" critique of contingent valuation. *Applied Economic Perspectives and Policy*, p.ppt029.

Hanemann, W.M., 1991. Willingness to pay and willingness to accept: how much can they differ?. *The American Economic Review*, 81(3), pp.635-647.

Hauser, J.R., Eggers, F., and Selove, M. (2016), *The Strategic Implications of Precision in Conjoint Analysis*, MIT Sloan School of Management working paper.

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In *The economics of new goods* (pp. 207-248). University of Chicago Press.

Hausman, J., 2012. Contingent valuation: from dubious to hopeless. *The Journal of Economic Perspectives*, 26(4), pp.43-56.

Jones, C. I., & Klenow, P. J. (2016). Beyond GDP? Welfare across countries and time. *The American Economic Review*, 106(9), 2426-2457.

- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of political Economy*, 98(6), 1325-1348.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The journal of economic perspectives*, 5(1), 193-206.
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
- Kuznets, S. (1934) "National Income, 1929-1932" 73rd Congress, 2nd Session, Senate Document no. 124.
- Kuznets, S. (1973). Modern economic growth: findings and reflections. *The American economic review*, 63(3), 247-258.
- Landefeld, J. S. (2000). GDP: One of the Great Inventions of the 20th Century. *Survey of Current Business*, 80(1), 6-14.
- List, J.A. and Gallet, C.A., 2001. What experimental protocol influence disparities between actual and hypothetical stated values?. *Environmental and Resource Economics*, 20(3), pp.241-254.
- Louviere, J.J., Hensher, D.A. and Swait, J.D., 2000. *Stated choice methods: analysis and applications*. Cambridge University Press.
- Manski, C.F., 1977. The structure of random utility models. *Theory and decision*, 8(3), pp.229-254.
- Marley, A. A. J. and Louviere, J. J., 2005. Some probabilistic models of Best, Worst, and Best-Worst choices. *Journal of Mathematical Psychology*. 49, 464-480.
- McFadden, D.L. (1974), 'Conditional logit analysis of qualitative choice behavior', in P. Zarembka (ed.), *Frontiers in Econometrics*, New York: Academic Press.

McFadden, D.L. (2014), In the Matter of Determination of Rates and Terms for Digital Performance in Sound Recordings and Ephemeral Recordings (WEB IV), Before the Copyright Royalty Board Library of Congress, Washington DC, Docket No. 14-CRB-0001-WR, October 6.

Miller, K.M., Hofstetter, R., Krohmer, H. and Zhang, Z.J., 2011. How should consumers' willingness to pay be measured? An empirical comparison of state-of-the-art approaches. *Journal of Marketing Research*, 48(1), pp.172-184.

Miller, M. (2012). How Google Made \$37.9 Billion in 2011. *Search Engine Watch*. (Accessed at: <https://searchenginewatch.com/sew/news/2140712/google-usd379-billion-2011>)

Morwitz, V.G., Steckel, J.H. and Gupta, A., 2007. When do purchase intentions predict sales?. *International Journal of Forecasting*, 23(3), pp.347-364.

Murphy, J.J., Allen, P.G., Stevens, T.H., Weatherhead, D., 2005. A meta-analysis of hypothetical bias in stated preference valuation, *Environmental and Resource Economics*, 30(3), pp.313-325.

Nakamura LI, and Soloveichik R (2015) Valuing 'free' media across countries in GDP. Federal Reserve Board of Philadelphia Working Paper No. 15-25.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2), 307-342.

Nordhaus, W. D. (2005). Schumpeterian Profits and the Alchemist Fallacy. *Yale Economic Applications and Policy Discussion Paper*, (6).

OECD (2008). *OECD compendium of productivity indicators*. (Accessed at: <http://www.oecd.org/std/productivity-stats/40605524.pdf>)

Peanut Labs (2015). *Peanut Labs Panel Book*. (Accessed at: http://news.peanutlabs.com/rs/peanutlabs/images/PL-PanelBook_2015.pdf)

Pepitone, J. 2012. Encyclopedia Britannica to stop printing books. CNN. (Accessed at: <http://money.cnn.com/2012/03/13/technology/encyclopedia-britannica-books>)

Plott, C.R. and Zeiler, K., 2005. The Willingness to Pay–Willingness to Accept Gap, the. *The American Economic Review*, 95(3), pp.530-545.

Rao, V.R., 2014. *Applied conjoint analysis* (p. 56). New York, NY: Springer.

Rosston, G., Savage, S. and Waldman, D., 2011. Household demand for broadband internet service. *Communications of the ACM*, 54(2), pp.29-31.

Schmalensee, R. (1978). Entry deterrence in the ready-to-eat breakfast cereal industry. *The Bell Journal of Economics*, 305-327.

Spence, M., B. Owen. 1977. Television programming, monopolistic competition, and welfare. *Q. J. Econ.* 103–126.

Stephens-Davidowitz, S., H. Varian. 2015. *A Hands-on Guide to Google Data*. (Accessed at: <http://people.ischool.berkeley.edu/~hal/Papers/2015/primer.pdf>)

Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2010). *Report by the commission on the measurement of economic performance and social progress*. Paris: Commission on the Measurement of Economic Performance and Social Progress.

Syverson, C. (2016). *Challenges to Mismeasurement Explanations for the US Productivity Slowdown* (No. w21974). National Bureau of Economic Research.

Thaler, R. H. (2015). *Misbehaving: The making of behavioral economics*. WW Norton & Company.

Thurstone, L.L., 1927. A law of comparative judgment. *Psychological review*, 34(4), p.273.

Varian H (2011, March 29) *The economic value of Google*. Presentation, San Francisco.

Varian, H. (Sep 2016). A microeconomist looks at productivity: A view from the valley. Presentation, Brookings (Accessed at: <https://www.brookings.edu/wp-content/uploads/2016/08/varian.pdf>)

Vickrey, W. (1961). Counterspeculation, Auctions and Competitive Sealed Tenders. *Journal of Finance*, 16, 8–37.

Völckner, F., 2006. An empirical comparison of methods for measuring consumers' willingness to pay. *Marketing Letters*, 17(2), pp.137-149.

Waldfoegel, J. (2012). Copyright Protection, Technological Change, and the Quality of New Products: Evidence from Recorded Music since Napster. *Journal of Law and Economics*, 55(4), 715-740.

Wertenbroch, K., & Skiera, B. (2002). Measuring Consumers' Willingness to Pay at the Point of Purchase. *Journal of Marketing Research*, 39, 228–241.

Whitehead, J.C., 2002. Incentive incompatibility and starting-point bias in iterative valuation questions. *Land Economics*, 78(2), pp.285-297.

Williams, B. 2008. “ A Hedonic Model for Internet Access Service in the Consumer Price Index.” *Monthly Labor Review*, July, pp.33-48.

Wlömert, N. and Eggers, F., 2016. Predicting new service adoption with conjoint analysis: external validity of BDM-based incentive-aligned and dual-response choice designs. *Marketing Letters*, 27(1), pp.195-210.

Tables and Figures

Table 1: Estimation results of binary logit model comparing valuation of Facebook in 2016 and 2017

	beta	Std. Error	z	p
(Intercept)	1.200	0.125	9.624	<0.001
log(E)	-0.309	0.030	-10.327	<0.001
Year_2017	0.290	0.209	1.385	0.166
Year_2017*log(E)	-0.101	0.051	-1.966	0.049

Table 2: Facebook value diagnostic

	beta	Std. Error	z	p
(Intercept)	0.321	0.254	1.261	0.207
log(E)	-0.346	0.032	-10.801	<0.001
Year_2017	0.306	0.220	1.392	0.164
Year_2017*log(E)	-0.105	0.054	-1.940	0.052
Facebook usage per week (scale)	0.117	0.043	2.740	0.006
Facebook number of friends (scale)	0.074	0.033	2.257	0.024
Facebook activity: Posting status updates or sharing pictures and videos (scale)	0.095	0.037	2.577	0.010
Facebook activity: Liking and commenting (scale)	0.093	0.039	2.363	0.018
Facebook activity: Playing games (scale)	0.054	0.024	2.234	0.025
Facebook activity: Using the messenger or chat (scale)	0.053	0.032	1.643	0.100
Facebook activity: Watching videos (scale)	0.066	0.037	1.748	0.080
Instagram user	-0.225	0.100	-2.245	0.025
Skype user	-0.067	0.092	-0.733	0.464
Google maps user	-0.076	0.107	-0.712	0.477
Google search user	-0.188	0.127	-1.482	0.138
YouTube user	-0.420	0.141	-2.983	0.003
Wikipedia user	0.049	0.096	0.510	0.610
Gender female (reference)	(0.000)			
Gender male	-0.220	0.086	-2.546	0.011
Age 18-24 (reference level)	(0.000)			
Age 25-34	-0.012	0.152	-0.079	0.937
Age 35-44	0.245	0.151	1.620	0.105
Age 45-54	0.367	0.155	2.371	0.018
Age 55-64	0.590	0.161	3.669	<0.001
Age 65+	0.936	0.176	5.335	<0.001

Income less than 25K (reference)	(0.000)			
Income 25K to 50K	0.081	0.140	0.578	0.563
Income 50K to 100K	-0.030	0.131	-0.229	0.819
Income 100K to 150K	-0.370	0.157	-2.355	0.019
Income 150K or more	0.441	0.165	2.671	0.008
Income "prefer not to answer"	0.784	0.273	2.873	0.004

Table 3: Estimation results of binary logit model comparing IC and NIC scenarios (2016 study)

	beta	Std. Error	z	p
(Intercept)	1.178	0.135	8.726	<0.001
log(E)	-0.449	0.034	-13.147	<0.001
IC	0.022	0.184	0.119	0.905
IC*log(E)	0.140	0.045	3.076	0.002

Table 4: Effect of random answers

Random sample R	Non-random sample	Mean intercept	Mean beta log (E)	Std. error Intercept	Std. error beta log(E)	WTA	Surplus	Scale S
800	700	0.517	-0.135	0.139	0.033	\$46.29	\$430.53	0.431
400	1100	0.846	-0.218	0.149	0.035	\$48.52	\$390.61	0.700
200	1300	1.020	-0.262	0.151	0.037	\$49.06	\$371.32	0.844
100	1400	1.122	-0.289	0.157	0.038	\$48.91	\$359.69	0.929
0	1500	1.206	-0.311	0.163	0.039	\$48.18	\$349.72	(1.000)

Table 5: Effect of sample size

Sample size	Mean intercept	Mean beta log (E)	Std. error Intercept	Std. error beta log(E)	mean WTA	95% CI lower	95% CI upper
200	1.242	-0.319	0.462	0.110	\$49.65	\$13.13	\$187.73
400	1.227	-0.316	0.324	0.077	\$48.72	\$21.16	\$112.28
800	1.214	-0.311	0.226	0.053	\$49.30	\$27.83	\$87.27
1500	1.206	-0.311	0.163	0.039	\$48.18	\$31.69	\$73.26

Table 6: Estimation results for the marginal effect of time (IC study)

	beta	Std. Error	z	p
(Intercept)	0.324	0.126	2.572	0.010
log(E)	-0.237	0.024	-10.009	<0.001
Time 1 week (reference)	(0.000)			
Time 2 weeks	0.235	0.135	1.734	0.083
Time 1 month	0.357	0.133	2.688	0.007

Table 7: Median WTA Estimates

Category	WTA/year 2016	WTA/year 2017	95% CI 2016		95% CI 2017		n
			lower	upper	lower	upper	
All Search Engines	\$14,760	\$17,530	\$11,211	\$19,332	\$13,947	\$22,080	8,074
All Email	\$6,139	\$8,414	\$4,844	\$7,898	\$6,886	\$10,218	9,102
All Maps	\$2,693	\$3,648	\$1,897	\$3,930	\$2,687	\$5,051	7,515
All Video	\$991	\$1,173	\$813	\$1,203	\$940	\$1,490	11,092
All E-Commerce	\$634	\$842	\$540	\$751	\$700	\$1,020	11,051
All Social Media	\$205	\$322	\$156	\$272	\$240	\$432	6,023
All Messaging	\$135	\$155	\$98	\$186	\$114	\$210	6,076
All Music	\$140	\$168	\$112	\$173	\$129	\$217	6,007

Figure 1: Share of Information sector's contribution to GDP (Source: BEA)

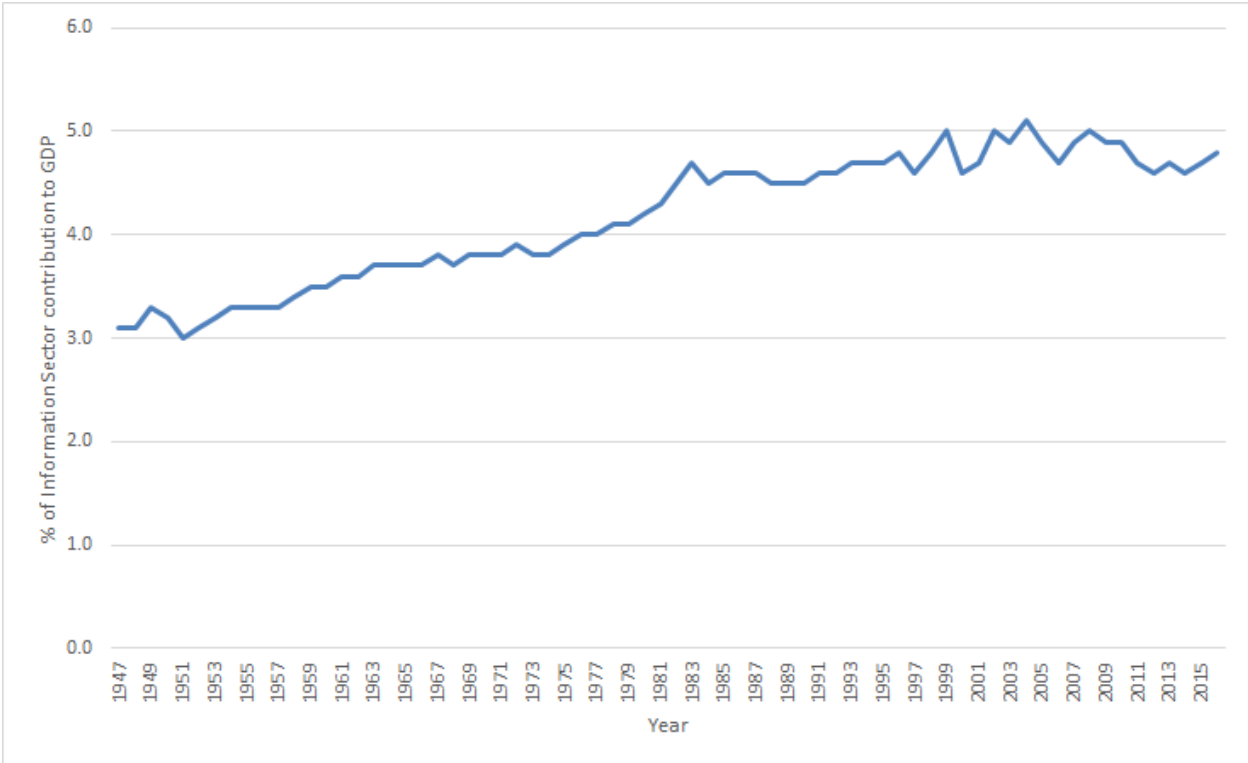


Figure 2: Consumer surplus and revenue for classic goods such as cars

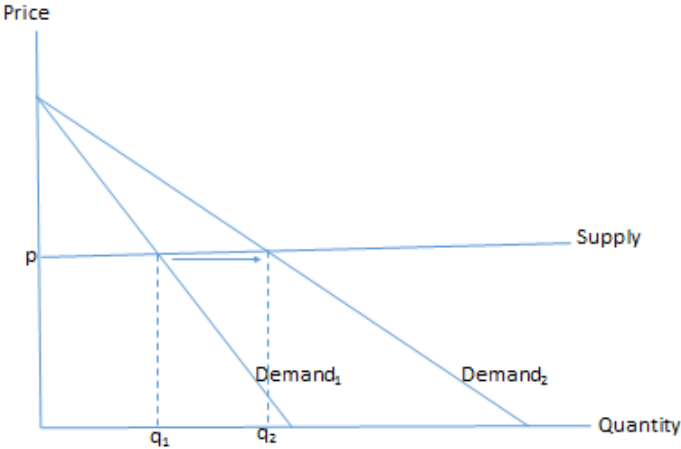


Figure 3: Consumer surplus and revenue for purely digital goods

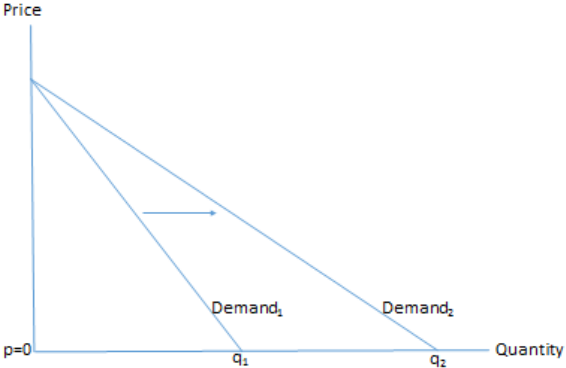


Figure 4: Consumer surplus and revenue for transition goods such as encyclopedias

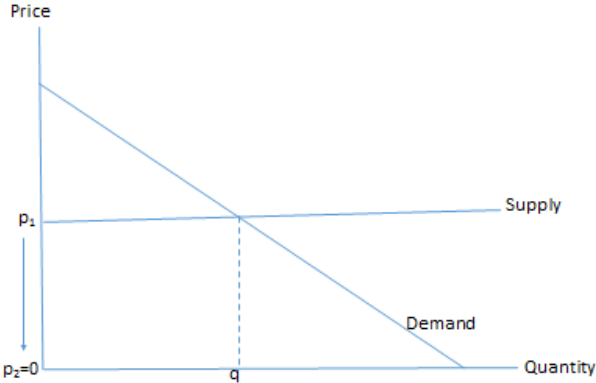


Figure 5: WTA demand curves for Facebook in 2016 and 2017

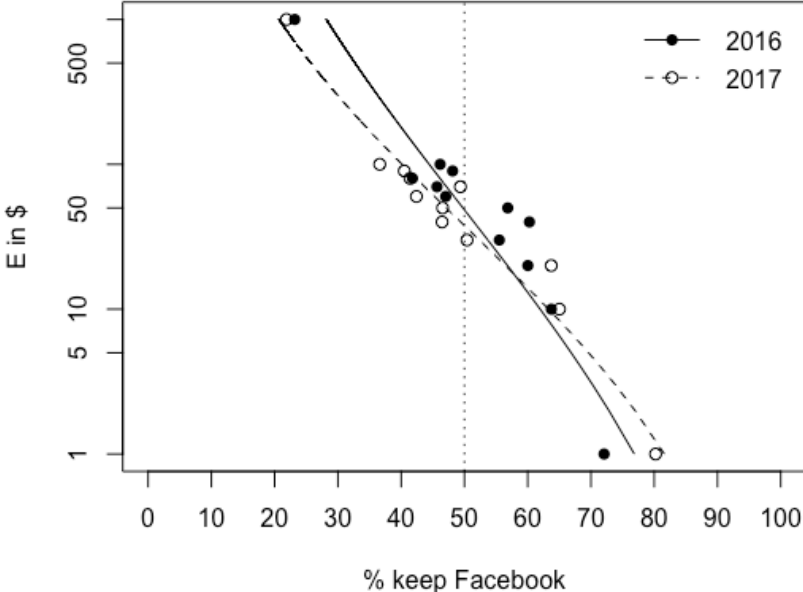


Figure 6: Comparison of BDM lottery and SBDC experiment

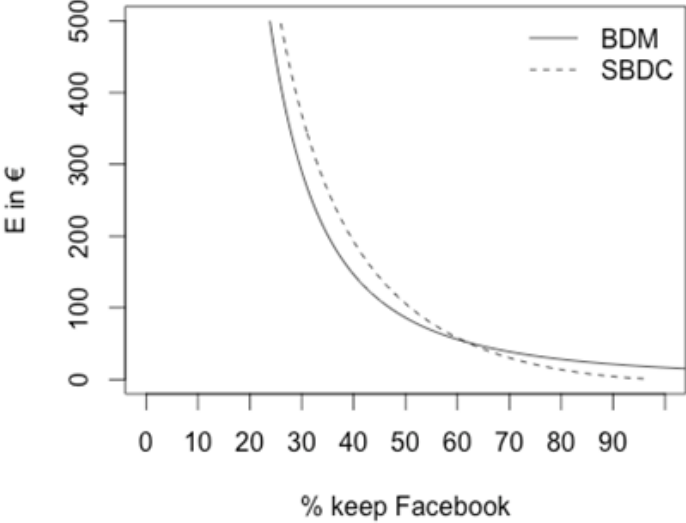


Figure 7: Assessment of hypothetical bias for Facebook

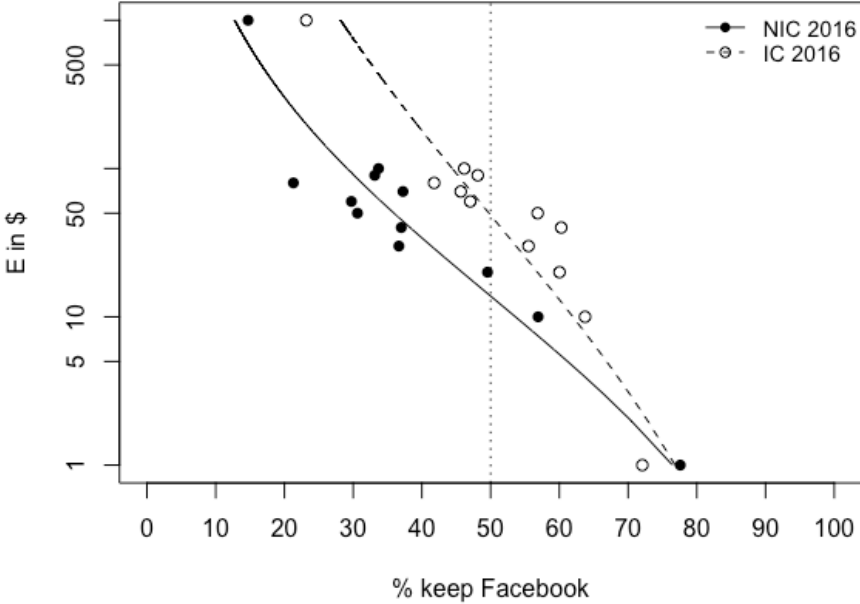


Figure 8: Effect of scale of the estimates on logit function

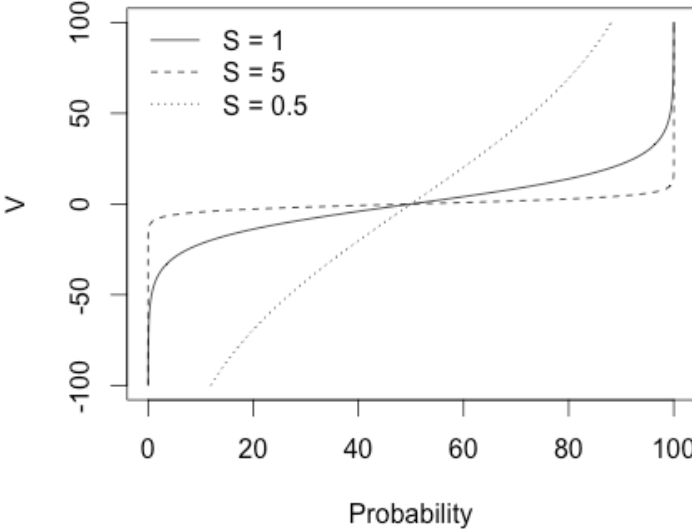


Figure 9: Effect of time on the probability to keep Facebook

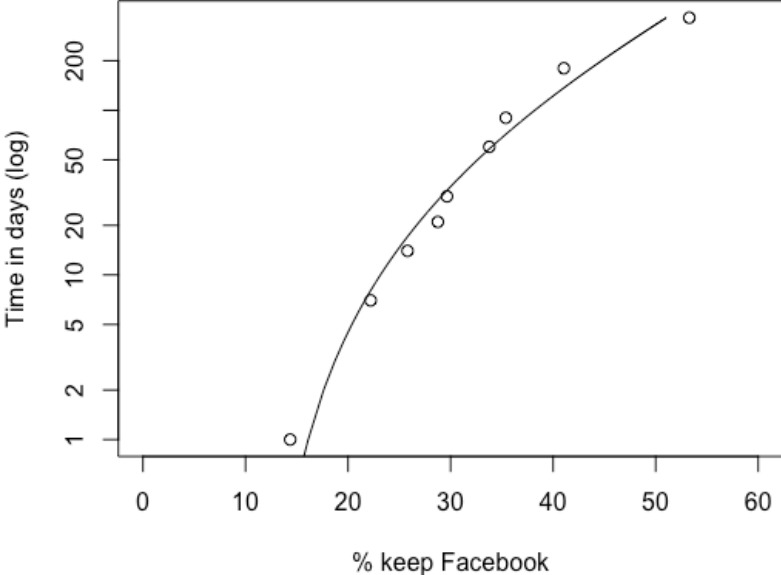


Figure 10: Example of Google Consumer Surveys

The Daily Globe


Top Stories World US Business Entertainment Sports search

Fair Use Digital Circulation Strategy Information Overload

The Work of Art in the Age of Mechanical Reproduction

Matthew Dodd from the January 16, 2013 issue

Jurgen Habermas R&D Android cops beat The Weekender mathewi Tim Carmody attracting young readers tweets, collaboration tags the medium is the message blog plagiarism horse-race coverage advertising the other longer Book Review....



Privacy put the paper to bed Fuego news.me photo source: proimos/flickr

Popular on The Daily Globe

MOST EMAILED MOST VIEWED

OPINION
[A Line in the Sand Against Rate Hikes](#)

ENTERTAINMENT
[A-List Guide to Oscar Parties](#)

THE BUZZ
[Memorable Quotes from 2012](#)

SPORTS VIDEO
[LeBron James Youngest to Score 20000](#)

HOME & GARDEN
[How to Plant a Winter Garden](#)

Please complete the following survey to access this premium content.

Would you prefer to keep access to Facebook or go without access to Facebook for 1 month and get paid \$200?

Keep access to Facebook

Give up Facebook and get paid \$200

OR

Show me a different question

Skip survey

Google [INFO](#) [PRIVACY](#)

Figure 11: WTA demand curves comparing 2016 (solid line) and 2017 (dashed line) for most widely used categories of digital goods

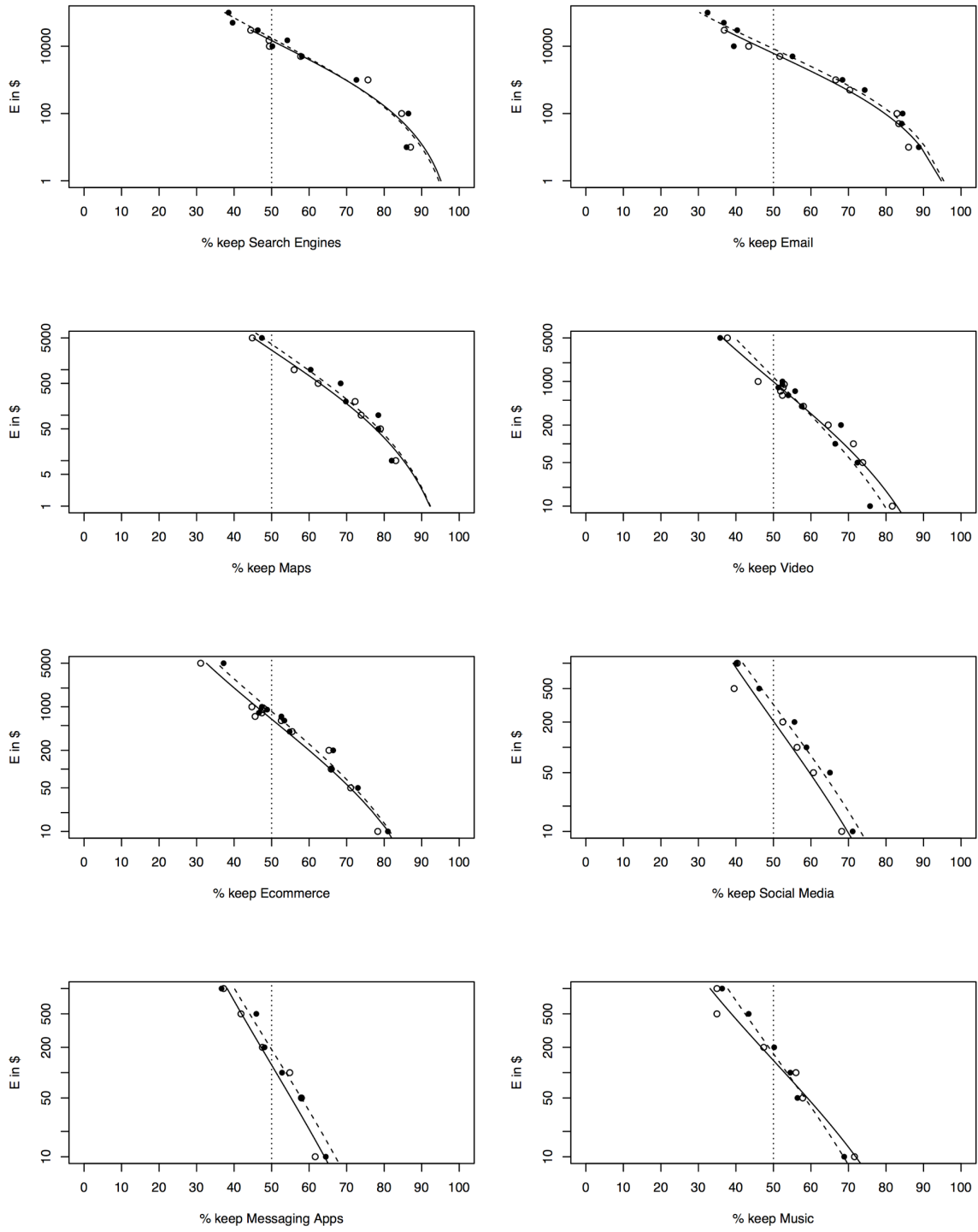


Figure 12: WTA demand curves for breakfast cereal

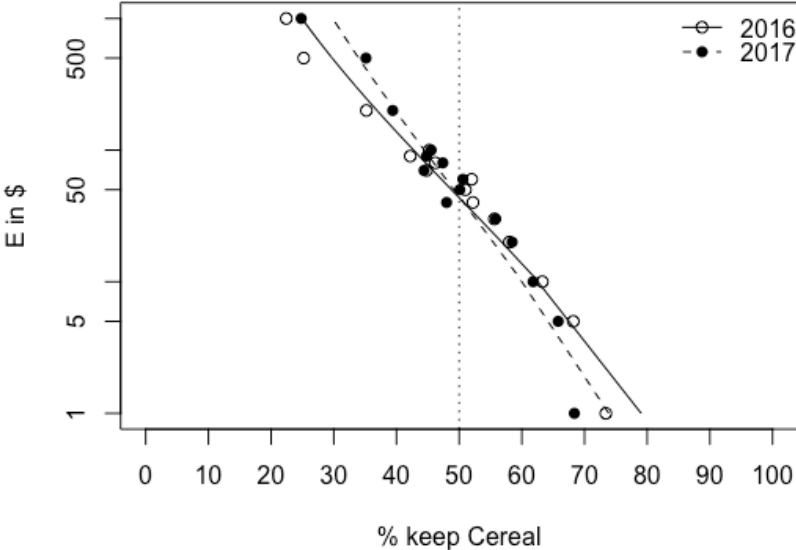
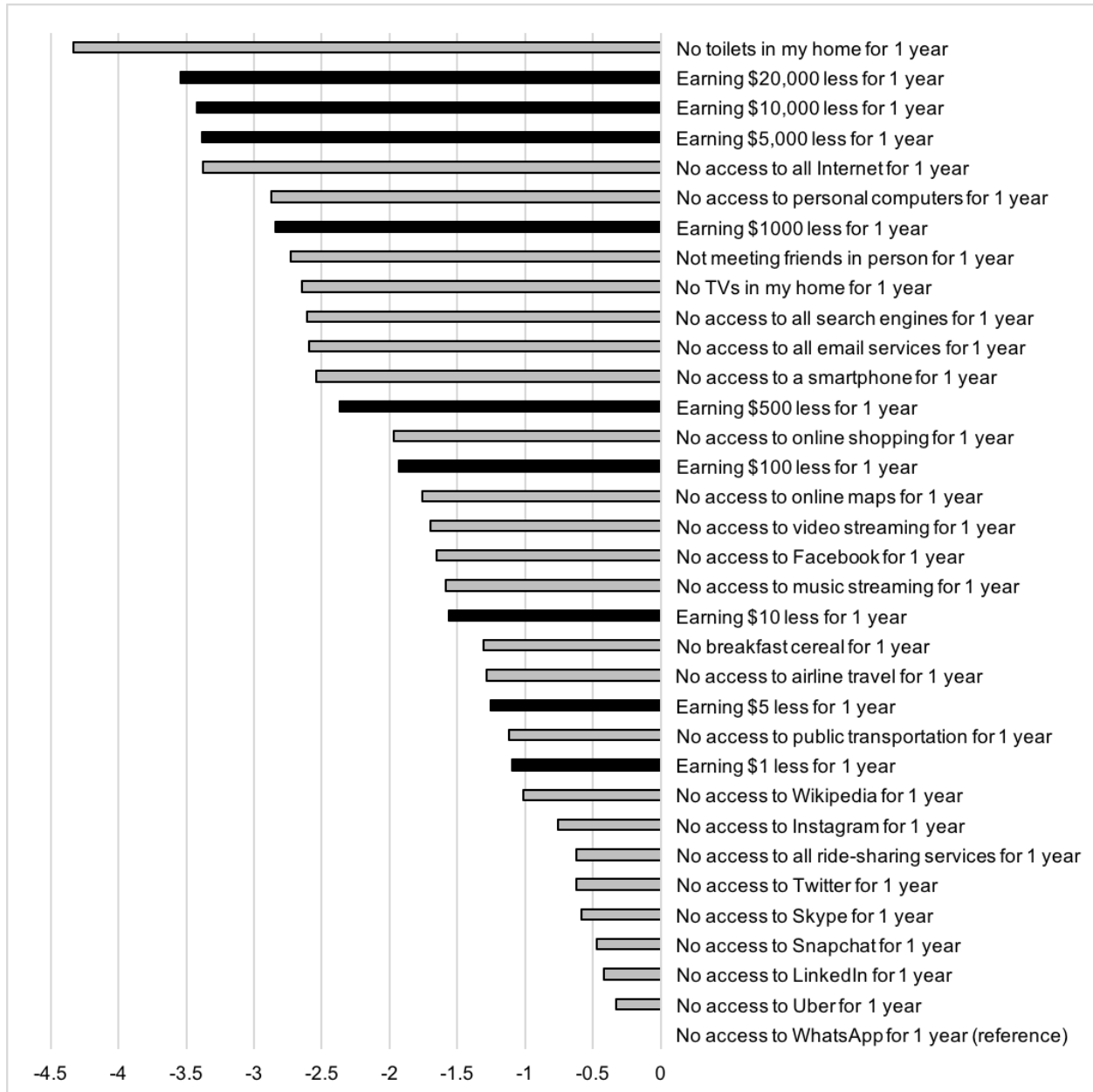


Figure 13: Exemplary best-worst scaling task

Please assume that you would have to give up access to some services or amenities for 1 year. Please consider the options below. Which of these options do you find worst and best?

	Option 1	Option 2	Option 3
	No breakfast cereal for 1 year	No access to online shopping for 1 year	Earning \$500 less for 1 year
Worst option:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Best option:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 14: (Dis-)Utility according to best-worst scaling



Appendix

Table A.1: Full estimation model for Facebook study

	beta	Std. Error	z	p
(Intercept)	1.178	0.135	8.726	<0.001
log(E)	-0.449	0.034	-13.147	<0.001
IC	0.022	0.184	0.119	0.905
IC*log(E)	0.140	0.045	3.076	0.002
Year_2017	-0.097	0.208	-0.465	0.642
Year_2017*log(E)	-0.039	0.054	-0.721	0.471
Year_2017*IC	0.386	0.295	1.310	0.190
IC*Year_2017*log(E)	-0.062	0.074	-0.838	0.402

Table A.2: Effect of time frame

	beta	Std. Error	z	p
(Intercept)	-1.650	0.060	-27.550	<0.001
log(T)	0.137	0.021	6.419	<0.001
log(T)^2	0.025	0.005	5.520	<0.001

Table A.3: Estimation results of binary logit model comparing Peanut Labs (non-incentive compatible group) and GCS

	beta	Std. Error	z	p
(Intercept)	0.579	0.114	5.091	<0.001
log(E)	-0.374	0.029	-12.686	<0.001
GCS	0.002	0.168	0.011	0.991
GCS*log(E)	0.031	0.043	0.715	0.474

Table A.4: Estimated logistic functions for most widely used categories of digital goods

	E-commerce	Email	Maps	Messaging	Music	Search	Social	Video
(Intercept)	2.108***	2.49***	2.12***	1.555***	1.669***	2.587***	1.827***	2.626***
log(E)	-0.351***	-0.342***	-0.316***	-0.234***	-0.362***	-0.313***	-0.282***	-0.345***
Year_2017	-0.013	0.033	0.028	0.054	-0.309.	-0.135	0.195	-0.279.
Year_2017*log(E)	0.014	-0.002	0.009	-0.003	0.084*	0.016	-0.012	0.057*
Age 18-24 (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Age 25-34	0.113	0.012	0.125	-0.257**	-0.022	-0.008	-0.175.	-0.092
Age 35-44	0.295***	0.096	0.339**	-0.2.	0.025	0.171.	0.001	-0.181*
Age 45-54	0.359***	0.472***	0.309**	-0.254*	-0.174	0.159	0.096	-0.301***
Age 55-64	0.401***	0.684***	0.255*	-0.295**	-0.314**	0.382***	-0.119	-0.588***
Age 65+	0.282**	1.089***	0.053	-0.338**	-0.552***	0.518***	-0.078	-0.555***
Age Unknown	-0.035	0.195	-0.108	0.078	0.308	0.248	0.013	0.096
Gender Female (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Gender Male	-0.099*	-0.117*	-0.099.	-0.355***	-0.023	-0.204***	-0.486***	-0.03
Gender Unknown	0.192	0.095	0.14	-0.669***	-0.582**	-0.209	-0.55*	-0.509***
Income \$0-\$24,999 (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Income \$25,000-\$49,999	-0.073	0.101	0.092	0.105	0.051	0.293**	-0.038	-0.074

Income \$50,000-\$74,999	-0.025	0.297**	0.29**	-0.058	0.02	0.348***	-0.039	0.019
Income \$75,000-\$99,999	-0.018	0.143	0.405**	0.131	-0.083	0.479***	-0.134	0.004
Income \$100,000-\$149,999	-0.012	0.036	0.992***	-0.101	0.357	0.435*	0.051	0.046
Income \$150,000+	0.344	0.026	0.843.	-0.503	-0.17	0.888.	0.373	-0.183
Income Prefer not to say	0.046	-0.562*	-0.068	-0.018	-0.023	0.399.	-0.148	0.034
Income Unknown	-0.154	0.149	-0.093	-0.354	0.094	0.146	-0.517.	-0.276
Urban Density Rural (reference)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Urban Density Suburban	0.064	0.101	0.073	-0.051	0.197*	-0.007	0.112	0.072
Urban Density Urban	0.024	0.218**	0.141.	-0.029	0.37***	0.137.	0.015	0.094
Urban Density Unknown	-0.038	-0.067	0.227	0.269	0.17	0.343*	0.128	-0.273.

*** p < 0.001, ** p < 0.01, * p < 0.05, . < 0.1

Table A.5: Best-worst scaling estimation results

Good	Utility	Str. Error	WTP implied from demand function
No toilets in my home for 1 year	-4.331	0.139	\$346'345.39
Earning \$20,000 less for 1 year	-3.540	0.144	\$18'079.67
Earning \$10,000 less for 1 year	-3.424	0.123	\$11'729.62
Earning \$5,000 less for 1 year	-3.382	0.161	\$10'023.70
No access to all Internet for 1 year	-3.373	0.123	\$9'694.25
No access to personal computers for 1 year	-2.870	0.134	\$1'482.92
Earning \$1000 less for 1 year	-2.839	0.117	\$1'323.45
Not meeting friends in person for 1 year	-2.725	0.116	\$866.24
No TVs in my home for 1 year	-2.647	0.116	\$645.66
No access to all search engines for 1 year	-2.610	0.115	\$563.80
No access to all email services for 1 year	-2.592	0.115	\$525.43
No access to a smartphone for 1 year	-2.542	0.115	\$437.16
Earning \$500 less for 1 year	-2.371	0.114	\$230.40
No access to online shopping for 1 year	-1.967	0.113	\$51.13
Earning \$100 less for 1 year	-1.933	0.113	\$45.03
No access to online maps for 1 year	-1.756	0.113	\$23.24
No access to video streaming for 1 year	-1.695	0.112	\$18.56
No access to Facebook for 1 year	-1.654	0.112	\$15.91
No access to music streaming for 1 year	-1.587	0.112	\$12.36
Earning \$10 less for 1 year	-1.565	0.112	\$11.41
No breakfast cereal for 1 year	-1.307	0.113	\$4.36

No access to airline travel for 1 year	-1.287	0.112	\$4.04
Earning \$5 less for 1 year	-1.254	0.127	\$3.58
No access to public transportation for 1 year	-1.120	0.113	\$2.17
Earning \$1 less for 1 year	-1.097	0.128	\$1.99
No access to Wikipedia for 1 year	-1.016	0.112	\$1.47
No access to Instagram for 1 year	-0.754	0.114	\$0.55
No access to all ride-sharing services for 1 year	-0.621	0.115	\$0.34
No access to Twitter for 1 year	-0.621	0.114	\$0.34
No access to Skype for 1 year	-0.586	0.114	\$0.30
No access to Snapchat for 1 year	-0.474	0.116	\$0.19
No access to LinkedIn for 1 year	-0.415	0.115	\$0.16
No access to Uber for 1 year	-0.326	0.117	\$0.11
No access to WhatsApp for 1 year (reference)	0.000		\$0.03

Figure A.1: Example of Incentive Compatible (IC) Questionnaire for Facebook SBDC question
(for E = \$80)

Would you prefer to keep access to Facebook or go without access to Facebook for 1 month and get paid \$80?

We want to reward you for thinking carefully about this question. Therefore, we will randomly pick 1 out of every 200 respondents and her/his selection will be fulfilled:

- If you choose "Keep access to Facebook" you can keep using Facebook as before. However, you will not receive the \$80 in cash.
- If you choose "Give up Facebook and get paid \$80" you will receive the \$80 in cash, provided that you do not access Facebook for 1 month. Facebook collects the date and time when you have last used your account. Given your permission, we can access this time with an app (e.g., see [this link](#) for an example). In order to get your permission, we will contact you via email. You can revoke this permission at any time.

Therefore, it is in your best interest to think carefully about how valuable you find Facebook.

What is your decision: Would you prefer to keep access to Facebook or go without access to Facebook for 1 month and get paid \$80?

Keep access to Facebook

Give up Facebook and get paid \$80

Proceed to next page

Figure A.2: Assessment of selection bias

